Titre: Title:	How to support trust in complex systems
Auteur: Author:	Hasmik Atoyan
Date:	2006
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Atoyan, H. (2006). How to support trust in complex systems [Master's thesis, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/7862/

UNIVERSITÉ DE MONTRÉAL

HOW TO SUPPORT TRUST IN COMPLEX SYSTEMS

HASMIK ATOYAN
DÉPARTMEMENT DE MATHÉMATIQUES
ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION DU DIPLÔME
DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INDUSTRIEL)
JUIN 2006

Your file Votre référence
ISBN: 978-0-494-19279-5
Our file Notre référence
ISBN: 978-0-494-19279-5

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

HOW TO SUPPORT TRUST IN COMPLEX SYSTEMS

présenté par: ATOYAN Hasmik

en vue de l'obtention du diplôme de: Maîtrise ès science appliquées

a été dûment accepté par le jury d'examen constitué de:

M. DESMARAIS Michel, Ph.D., président

M. ROBERT Jean-Marc, Doctorat, membre et directeur de recherche

M. DUQUET Jean-Rémi, Ph.D., membre

AKNOWLEDGEMENTS

I thank Professor Jean-Marc Robert for his guidance of my research and constructive ideas and criticism that helped to develop a right scientific attitude to various problems of cognitive science reflected in this memoir.

My thanks also go to the members of R&D group of Lockheed Martin Canada, led by Dr Elisa Shahbazian, for their support and collaboration in this project. I am grateful in particular to Yanick Allard for his time and efforts in explaining the concepts and operational principles of his novel software IDFS that I have evaluated in this memoir, for his collaboration and fast implementation in the IDFS of the developed recommendations directed to the improvement of the interface usability related to the aspect of trust. I thank Jean-Rémi Duquet for his clear and helpful advices, numerous discussions on the concept of trust in complex systems. Thanks to Jean Couture for his help in correcting the French translation of this memoir.

I also would like to acknowledge the help of Professor François Cavayas in finding qualified users of remote sensing applications at the Department of Geography of the University of Montreal for the IDFS evaluation. At last, my special thanks go to all expert users and graduate students who agreed to participate in this evaluation.

RÉSUMÉ

Mots clefs: Confiance, mesure, dans les systèmes automatisés, interface humaine-machine, lignes directrices, utilisabilité de l'interface.

Un des principaux défis à relever au sujet de l'utilisation sécuritaire et efficace des nouveaux systèmes automatisés dans les systèmes complexes porte sur le niveau de confiance que les opérateurs développent à l'égard du système.

Ce mémoire vise deux objectifs. Le premier est de proposer des règles et des lignes directrices aux concepteurs de systèmes pour que les utilisateurs aient un niveau de confiance adéquat envers les systèmes automatisés, en particulier envers les nouveaux systèmes. Les règles et les lignes directrices sont basées sur une vaste revue des études théoriques, expérimentales et empiriques sur le sujet.

Le deuxième objectif est de tester l'impact de l'utilisabilité du système sur le niveau de la confiance à l'égard du système. Notre hypothèse est que l'amélioration de l'utilisabilité de l'interface favorise le développement de la confiance. Nous avons évalué l'utilisabilité de l'interface d'une nouvelle application de la télédétection, IDFS (Intelligent Data Fusion System) par la méthode heuristique (impliquant l'auteure de la présente étude) et l'évaluation pas à pas (menée avec six utilisateurs experts du domaine de la télédétection). Pendant l'évaluation, nous nous sommes particulièrement attardées aux problèmes pouvant avoir une incidence sur la confiance. Nous avons développé des recommandations pour améliorer l'utilisabilité de l'interface, et celles-ci ont été implantées par les développeurs de l'application. En particulier, nous avons amélioré la compatibilité du comportement du système par rapport aux attentes des utilisateurs, le retour de l'information, la clarté visuelle et la traçabilité de l'information.

Les deux interfaces, initiale et modifiée, ont été testées par six utilisateurs (étudiants gradués de l'Université de Montréal au département de géographie faisant la recherche dans le domaine de la télédétection). Pour mesurer la confiance dans le système on a utilisé l'échelle d'évaluation de Jian et al. (2000). Les résultats ont révélé que, bien que

les utilisateurs aient été conscients que les algorithmes et les fonctionnalités des deux interfaces étaient les mêmes, la confiance pour la nouvelle version de l'interface était presque deux fois supérieure à la version initiale. Les résultats montrent donc que l'utilisabilité sert à la confiance. Les résultats vont aussi dans le sens de la proposition selon laquelle au début de l'interaction avec un nouveau système, il s'agit fortement de confiance affective et de confiance analogique (la cohérence du comportement du système avec les règles et les procédures). On conclut que l'amélioration de la qualité de l'interface, en particulier la compatibilité, le retour de l'information, la clarté visuelle et la traçabilité de l'information, supportent le développement des confiances affective et analogique.

La recherche montre aussi que l'échelle d'évaluation de Jian et al. (2000) permet de mesurer la confiance initiale à l'égard du système (c.-à-d. la confiance pendant l'interaction avec un nouveau système). Toutefois, cette échelle ne donne pas de description de la terminologie utilisée. On propose d'expliquer chaque item afin de permettre aux utilisateurs de bien comprendre. Il faut également mentionner que le l'échelle fournit 12 réponses pour 12 différents items reliés, ne spécifie pas leur poids respectif et ne donne pas de mesure unique de la confiance globale à l'égard du système. Il serait pratique et informatif d'unifier les réponses de façon à pouvoir estimer la confiance globale à l'égard du système.

À partir de cette échelle d'évaluation, on propose un modèle permettant de mesurer la confiance globale envers le système. Pour les deux interfaces, la dispersion des valeurs des différents items de l'échelle par rapport à la valeur de la confiance globale est petite, $\sigma \approx 0.077$ pour l'ancienne et $\sigma \approx 0.076$ pour la nouvelle version de l'interface. Cela montre que les différents items sont très proches les uns des autres ou qu'ils mesurent des aspects de la confiance que les sujets ont de la difficulté à bien différencier.
L'étude présente aussi les limitations des résultats expérimentaux et du modèle proposé. Il aurait été préférable de mesurer la confiance sur un plus grand nombre des tâches et de fonctionnalités du système, et avec un plus grand nombre d'utilisateurs.

Dans le futur, les recherches pourront explorer et mettre en évidence ces différences. Par exemple, une plus grande dispersion peut être un indicateur d'une plus grande confusion par rapport au sens de l'item. Par contre, la dispersion ne peut pas être le seul critère pour assigner des poids. D'autres facteurs devraient être pris en considération, comme l'importance de chaque item pour la confiance.

Une autre piste de recherche serait de mesurer l'impact des autres qualités de l'utilisabilité comme l'aide et la guidage, la cohérence, la prévention des erreurs, la flexibilité et la rapidité de l'utilisation sur la confiance. Une façon de mesurer l'impact des différentes qualités de l'interface sur la confiance serait de réaliser une expérience avec deux versions de l'interface : l'une avec ces qualités de l'utilisabilité et l'autre sans ces qualités. La comparaison de ses résultats va montrer les qualités de l'interface qui jouent un plus grand rôle dans le développement de la confiance.

ABSTRACT

Keywords: Trust, automated systems, system misuse and disuse, human computer interface, guidelines, interface usability, trust measurement.

One of the major factors for safe and efficient utilisation of new automated complex systems is the level of trust that an operator has in the system.

This study encompasses two objectives. The first one is to propose design rules and guidelines for system designers on how to support an appropriate level of trust in automated systems, in particular in novel systems. The proposed guidelines are based on extensive theoretical, experimental and empirical results on trust in automated systems available in the research literature.

The second objective is to verify the impact of interface usability on the level of trust in the system. We hypothesized that the enhancement of the interface usability supports trust. In order to test this hypothesis the usability of a new remote sensing application IDFS (Intelligent Data Fusion System) was evaluated by two methods, one based on heuristics (conducted by the author of this study) and the other based on a walkthrough (conducted with six expert users in the domain of remote sensing) methods. During the evaluation a particular importance was paid to the issues of trust. We developed recommendations on possible improvements of usability of the IDFS interface aimed at trust enhancement. Then these recommendations were implemented in the modified version of IDFS. In particular, the compatibility of the system's behavior with the user's expectations, informative feedback, visual clarity and traceability of the information have been enhanced.

Both the new (redesigned) and the old (initial) interfaces of the IDFS have been tested by six users, namely, graduate students doing research in the domain of remote sensing at the University of Montreal (Department of Geography). The level of trust was

measured using the trust measurement scale of Jian et al. (2000). Even though the operators were well aware that the system functionalities in the two versions of the IDFS were mostly the same, the measurements revealed that the new interface did increase trust by a factor of two. Thus, these results support the hypothesis that usability promotes trust. They also are consistent with the proposition that in early interactions with a new system trust strongly depends on affective[1] and analogical[2] factors. It is thus concluded that enhancement of interface qualities such as compatibility of the system's behavior with the user's expectations, informative feedback, visual clarity and traceability, which all support development of the affective and analogical trust.

The test results also show that the scale of Jian et al. (2000) does allow the measurement of the initial trust in novel systems (i.e. when there is no prior experience with the system). However, the scale does not provide a description for the terminology that is used. We propose to explicitly explain each item of the scale in order to help the users understand the terminology. Another point is that the scale of Jian et al. (2000) consists of 12 answers for 12 different trust-related items without proposing unified single measure for the *overall* trust in the system.

In the study we built on the trust measurement scale of by Jian et al. (2000), and suggested a model that enables measurement of the overall trust. For both the old and the new interfaces the dispersion of individual items in the scale from the overall trust value is small: $\sigma_{old} \approx 0.077$ and $\sigma_{new} \approx 0.076$, respectively. This reveals that the individual items in this scale are very close to each other or they measure aspects of trust that is hard for the subjects to differentiate.

---

[1] Affective trust tuning involves essentially anything which results in producing positive or negative feelings toward the system.

[2] Analogical trust is developed by comparing the system behavior with rule-based expectations. When rules are consistent with trustworthy behavior, they can increase people's expectations of satisfactory performance.

The limitations of experimental results and the trust measurement model we have developed are also discussed. It would have been preferable to measure trust with a larger number of users, tasks, and functionalities in the system.

Future research is necessary to reveal these differences. For example, larger dispersion could be an indicating for a larger confusion in the meaning of item. However, dispersion cannot be the single criteria for assigning the weight. Other factors should be taken into account, such as direct bearing of that each item to trust.

Another future research avenue is to find out the impact of other usability qualities such as help and guidance, consistency, error prevention, flexibility and rapidity of use on trust. One way to measure the impact of different interface qualities on trust is to design an experiment with two versions of interface: one with the interface qualities and the other without them. The comparison of these results will show which interface qualities play major role in trust development.

CONDENSÉ EN FRANÇAIS

Comment soutenir la confiance dans les systèmes complexes

Un des principaux défis à relever au sujet de l'utilisation sécuritaire et efficace des nouvelles technologies dans les systèmes complexes porte sur le niveau de confiance que les opérateurs développent à l'égard du système. Deux problèmes peuvent affecter le niveau de confiance: la sous-confiance et la sur-confiance de l'opérateur. La sous-confiance est un bas niveau de confiance de l'opérateur envers le système et peut causer une mauvaise utilisation du système. Il se peut qu'un système, même fiable, ne soit pas utilisé par l'utilisateur si ce dernier ne fait pas assez confiance au système. La sur-confiance est une confiance excessive de l'opérateur envers le système. Cela peut aussi entraîner une mauvaise utilisation du système, en ce sens que l'opérateur peut ne pas détecter les fautes et les erreurs du système. Ces deux types de problèmes peuvent compromettre la sécurité et la rentabilité du système et avoir des conséquences graves. Afin d'éviter un niveau de confiance inapproprié, il est important d'obtenir une bonne correspondance entre le niveau de confiance de l'opérateur et les capacités du système.

Cette étude présente une vaste revue d'études théoriques, expérimentales et empiriques sur la confiance dans les systèmes automatiques. Son but est d'aider les concepteurs de systèmes en leur proposant des règles et des lignes directrices pour concevoir des systèmes qui vont amener les opérateurs à développer un niveau de confiance adéquat envers les nouveaux systèmes d'aide à la décision.

L'étude présente et analyse les outils existants pour mesurer la confiance. On présente l'échelle de Jian et al. (2000) qui permet de mesurer les différentes composantes de la confiance. À partir de cette échelle, on propose un modèle qui permet de mesurer la confiance globale envers le système.

Cette étude présente aussi les résultats expérimentaux relatifs à l'impact de l'utilisabilité de l'interface sur la confiance.

Les règles et les lignes directrices suggèrent qu'il y a certains liens entre l'utilisabilité de l'interface et la confiance envers le système

Définition et composantes de la confiance

Dans une interaction humain-machine, on définit la confiance dans un contexte complexe comme suit (Madson et Gregor, 2000):
«La confiance est le degré selon lequel l'utilisateur est prêt à agir d'après les recommandations, les actions et les décisions des outils, des systèmes ou de l'aide à la décision.»
La confiance est un concept multidimensionnel et dynamique. Le terme multidimensionnel signifie que la confiance dépend de différentes composantes. Les principales composantes identifiées dans la littérature scientifique sont: la fiabilité, la prédiction, la compréhension, la robustesse du système, la familiarité, la foi, la compétence, l'intégrité, l'explication de l'intention, l'utilité, la réputation et la rumeur.

Lee et Moray (1992) ont identifié trois éléments de base dans le développement de la confiance. Ce sont la performance, le processus et l'objectif.
La performance décrit ce que l'automatisation fait. Elle est basée sur les observations directes de l'opérateur. Cela inclut la fiabilité, la prédiction et la robustesse du système.
Le processus décrit la façon dont le système fonctionne. C'est la compréhension des mécanismes, des algorithmes sous-jacents au système. Cela inclut l'intégrité et la compréhension du système.
L'objectif décrit pourquoi l'automatisation a été développée. Cela reflète l'intention du concepteur du système.

Pendant les premières interactions avec un nouveau système, la confiance dépend de l'objectif et non pas de la performance, parce que l'opérateur a peu d'expérience avec le système, il a plutôt des information sur l'objectif du système. Ce n'est qu'après avoir

acquis de l'expérience avec ce système qu'il pourra développer une opinion à l'égard de la performance et du processus du système. Alors, c'est en fournissant à l'opérateur de l'information sur l'objectif, la performance et le processus du système qu'on l'aidera à développer la confiance envers ce système. Toutefois, cette information doit être présentée de façon cohérente avec les processus cognitifs sous-jacents au développement de la confiance.

Confiance et processus cognitifs

See et Lee (2004) affirment que les humains développent et ajustent leur confiance de trois manières différentes: analytique, analogique et affective.

Pendant *le processus analytique*, l'information est traitée et analysée d'après l'expérience, la connaissance et le modèle mental qu'on a du système. Ce processus demande les ressources cognitives les plus élevées.

Pendant *le processus analogique*, les jugements sont faits d'après une comparaison du système avec les règles et les procédures. Le processus analogique demande moins de ressources cognitives que le processus analytique. La confiance ainsi acquise peut être fragile pendant les situations anormales si le comportement du système n'est pas compatible avec les règles. Cela peut détruire la confiance. La confiance analogique peut être aussi basée sur les réputations (par exemple celle de la compagnie qui a développé le système) et l'opinion d'autrui (les autres utilisateurs font confiance au système).

Les aspects affectifs ont une influence très importante sur le développement de la confiance. Les émotions prévalent quand les règles ne fonctionnent plus et quand il n'y a pas de ressources cognitives disponibles pour analyser la situation. Le processus affectif demande les ressources cognitives les moins élevées. Les humains basent leur jugement sur les technologies non seulement sur ce qu'ils pensent, mais aussi sur ce qu'ils ressentent à propos de celles-ci (Alkahami and Slovic, 1994).

Au début de l'interaction avec un nouveau système, l'opérateur n'a pas d'expérience quant à la performance et au comportement du système. La seule information qu'il peut avoir pour faire ou ne pas faire confiance au système, c'est l'information émotionnelle. Si cette information est assez forte pour engendrer une forte méfiance au système, il peut arriver que l'opérateur ne recueille plus d'information et arrête d'utiliser l'outil ou le système. Mais si la confiance émotionnelle est à un niveau modéré, l'opérateur continuera d'utiliser le système et de développer la confiance analogique. Après avoir acquis un niveau modéré de confiance analogique, il développera la confiance analytique.

Une des façons de soutenir la confiance basée sur les processus affectif et analogique, c'est de fournir à l'opérateur une bonne étiquette de l'ordinateur. Des exemples de bonne étiquette sont : une terminologie et un style de communication familière. Les résultats expérimentaux révèlent qu'une mauvaise étiquette peut diminuer la performance et la confiance envers un système même avec une fiabilité enlevé.

La confiance et la fiabilité du système

Avant de commencer à discuter des relations entre la confiance et la fiabilité du système, il est important d'établir la différence entre ces deux termes. La confiance est un état psychologique et peut être estimé par une évaluation subjective. La fiabilité peut être estimée par l'évaluation subjective et par la performance. On peut se fier à l'automatisation même si la confiance est faible: par exemple, l'utilisateur peut savoir que l'automatisation a des lacunes mais, néanmoins se fier au système à cause d'une surcharge du travail ou bien parce qu'il n'y a pas d'autre façon d'accomplir la tâche.

La confiance peut être affectée par la fiabilité du système. La fiabilité est la probabilité que le système ne fera pas d'erreurs. Une fiabilité élevée permet à l'opérateur d'observer le système et de développer la confiance.

La confiance dépend du niveau de l'automatisation. Parasuraman et al. (2000) identifient quatre niveaux d'automatisation : acquisition de l'information, analyse de l'information, sélection de la décision et de l'action, implémentation de l'action. Lorsque les données brutes sont disponibles, l'opérateur peut porter son attention alternativement aux données brutes et à l'automatisation. Il peut se fier au système et l'utiliser même si sa confiance à l'égard du système n'est pas élevée. Cela est possible pour les deux premiers niveaux d'automatisation mais pour les autres niveaux, et en particulier le quatrième niveau, la confiance envers le système est nécessaire pour qu'il soit utilisé.

Les résultats expérimentaux (Madhavan et Wiegman, 2005) montrent que les situations où le système est opaque (c.-à-d. permet très peu ou pas du tout d'accès aux données brutes) et où la fiabilité du système est faible sont propices au développement du problème d'ancrage cognitif. L'ancrage cognitif est la tendance de l'humain à être attaché à une hypothèse initiale qu'il a sélectionnée. Dans le système opaque, l'opérateur peut deviner la cause de l'état du système et prendre les décisions seulement d'après cette supposition. Cet effet d'ancrage cognitif peut augmenter la probabilité que l'opérateur ne soit pas d'accord avec les résultats du système, ce qui peut occasionner une méfiance accrue à l'égard du système. Une des conséquences possibles est alors que le système ne soit pas utilisé par l'opérateur.

La confiance peut être fortement affectée par les défauts de l'automatisation. Les résultats expérimentaux révèlent que les opérateurs s'attendent à des résultats parfaits de l'automatisation tandis que leurs attentes sont beaucoup moins grandes par rapport aux humains. À cause de ces attentes, les défauts de l'automatisation déclenchent un déclin rapide de la confiance. Les résultats expérimentaux montrent aussi que les humains font confiance et se fient à une automatisation imparfaite s'ils sont conscients de cette imperfection.

D'autres résultats montrent que les défauts dans les tâches faciles diminuent davantage la confiance que les défauts dans les tâches difficiles. Un autre résultat intéressant (Muir

and Moray, 1996) est que la méfiance dans une fonction particulière peut s'étendre aux autres fonctions du même sous-système. Cela peut mener à une méfiance injustifiée. Cependant, la méfiance ne s'étend pas aux systèmes similaires s'ils sont indépendants.

L'expérience initiale a un effet durable sur la confiance: si la confiance commence à un bas niveau, le système peut être mal utilisé, ce qui peut entraîner un niveau encore plus bas de confiance. La confiance est donc plus durable si elle est élevée au départ.

La confiance selon différents contextes

La confiance dépend du contexte. Il est important de comprendre comment les paramètres du contexte influent sur la confiance. Par exemple, Muir (1987) a trouvé que les gens font plus confiance aux systèmes automatisés dans les situations peu familières, mal connues. Il a montré (1994) que le risque agit aussi sur la confiance. L'opérateur pèse chaque cas différemment dépendant du risque impliqué. Les situations ayant un risque élevé demandent plus de confiance et l'opérateur a besoin de plus de temps de récupération lorsque des défauts d'automatisation surviennent.

La confiance peut être affectée par la confiance en soi (Lee and Moray, 1994). Si celle-ci est faible, l'opérateur est plus enclin à faire confiance à l'automatisation.

La confiance est aussi influencée par les situations où les opérateurs sont responsables de plusieurs tâches ou ont une surcharge élevée de travail. Ces situations lorsque combinées avec une fiabilité d'automatisation élevée, peuvent créer une confiance excessive envers le système. Cela peut entraîner un contrôle moins serré et donc une moins bonne détection des défauts du système par l'opérateur (Parasuraman et al, 1993).

L'utilisation de l'automatisation adaptative peut réduire la sur-confiance dans l'automatisation et augmenter la détection des défauts (Parasuraman et al., 1996). L'automatisation adaptive entraîne une série de transitions entre le contrôle manuel et automatisé. Elle contribue à empêcher la dégradation des habiletés et de la conscience de

la situation de la part des opérateurs (Parasuraman et al., 2000; Miller et al., 2005) parce que l'opérateur participe au processus.

Le stress et la confiance sont aussi interreliés. Plus le stress est grand plus il devient difficile de prendre des décisions et plus les différences individuelles et les processus cognitive comme les stéréotypes ressortent (Nickerson et Reily, 2004, Jiang et al., 2004). Cela aussi entraîne une confiance inadéquate.

Le contexte environnemental influence aussi la confiance. Bisantz et Seong (2001) ont trouvé que la confiance peut être affectée différemment par la source des défauts d'automatisation. La confiance sera ainsi davantage affectée par les défauts internes du système automatisé (par ex., un bogue du logiciel) que par les défauts externes, c.-à-d. ceux qui ne proviennent pas du système automatisé (par ex., sabotage, problèmes avec le transformateur).

Il est aussi important de mentionner les contextes humains, organisationnels et culturels (Lee et See, 2004). Les gens qui sont plus croyants, qui ont tendance à avoir davantage confiance aux autres, vont vouloir déléguer plus facilement les tâches au système automatisé. Le contexte organisationnel reflète l'interaction entre les gens. Cela inclut en particulier la réputation et les rumeurs. Le contexte culturel influe sur la confiance par les normes et les attentes sociales.

La confiance et le contenu et formatage de l'affichage

Le contenu et le formatage de l'affichage ont aussi un impact sur le développement de la confiance. L'information doit être organisée de manière cohérente en fonction des exigences de la tâche et des attentes des utilisateurs. L'affichage de l'information doit soutenir les processus affectif, analogique et analytique du développement de la confiance. Si l'information n'est pas disponible ou si elle est formatée d'une manière inadéquate, la confiance ne se développera pas correctement.

Souvent, la confiance et la crédibilité dépendent des caractéristiques de l'interface qui n'ont aucun lien évident avec les capacités du système (Briggs et al., 1998). Par exemple, Kim et Moon (1998) ont découvert que des caractéristiques comme des couleurs froides et pastel ou une disposition bien balancée, augmentent la perception de la confiance dans l'information.

Les outils pour mesurer la confiance.

La confiance dans le système peut être mesurée au moyen de méthodes objectives et subjectives (EATMP, 2003). Une méthode simple d'obtenir des mesures objectives est de mesurer la fréquence de l'utilisation du système ou d'outils du système. Il peut malgré tout arriver que l'opérateur fasse confiance au système mais ne l'utilise pas pour une autre raison. L'inverse est aussi vrai, l'opérateur peut tout de même utiliser le système même s'il a peu confiance en lui, ne serait-ce que parce qu'il n'y a aucune autre possibilité d'effectuer la tâche. Wickens et Xu (2002) considèrent que la confiance est un état purement psychologique et qu'elle peut être mesurée seulement par une évaluation subjective. Il semble que les mesures subjectives basées sur les réponses des opérateurs représentent l'approche la plus appropriée (EATMP, 2003).

Pour mesurer la confiance, des échelles d'évaluation multiples sont utilisées, par exemple celles de Muir et Moray (1996) Madson et Gregor (2000) et Jian et al. (2000). Les deux dernières échelles sont basées sur des résultats empiriques. Madson et Gregor (2000) ont développé une échelle pour mesurer la confiance dans les systèmes intelligents, l'échelle HCT (Human Computer Trust). Cette échelle se compose de cinq éléments principaux où chaque élément inclut cinq sous-éléments. Cette échelle considère que les opérateurs ont déjà plusieurs mois d'expérience avec le système, ce qui constitue une des limitations de cette échelle.

Jian et al. (2000) ont développé la première preuve que les concepts de confiance et de méfiance peuvent être traités comme des extrêmes sur le même continuum, c.-à-d. que

les deux concepts peuvent être mesurés sur la même échelle d'évaluation. Leur une échelle qui inclut 12 items et 7 degrés d'appréciation (allant de "pas du tout'' à ''extrêmement''). Toutefois, l'échelle d'évaluation fournit 12 réponses pour 12 différentes composantes reliées à la confiance et ne donne pas une réponse unique pour la confiance globale à l'égard du système. Il serait pratique d'unifier les réponses pour estimer la confiance globale à l'égard du système. À partir de cette échelle d'évaluation, on propose un modèle permettant de mesurer la confiance globale envers le système.

Il faut souligner qu'un grand soin doit être apporté aux mots utilisés dans l'échelle de l'évaluation, en particulier lorsque la langue maternelle des utilisateurs n'est pas l'anglais. Des termes comme 'reliability' et 'trust' peuvent avoir différentes significations pour des opérateurs de différente nationalité (EATMP, 2003).

Règles et lignes directrices

À partir d'une revue d'études théoriques, expérimentales et empiriques, nous proposons des règles et des lignes directrices pour concevoir des systèmes qui vont soutenir le développement d'un niveau de confiance adéquat chez les utilisateurs des nouveaux systèmes automatisés. Ces règles et les lignes directrices sont les suivantes :

Règles générales de la conception

1. **Concevoir le système pour un niveau de confiance approprié, ni trop élevé, ni trop bas.** Les deux situations, la sous-confiance et la sur confiance, peuvent compromettre la sécurité et la rentabilité du système (Parasuraman & Riley, 1997; Parasuraman & Miller, 2004; Lee and Sanquist, 2000; National Transportation Safety board, 1997; Zuboff, 1988).
2. **Pendant la période initiale, la période d'introduction au système, il faut préparer le système (le matériel et la logiciel) et l'utilisateur à certaines erreurs possibles ou fautes du système.** L'expérience initiale a un effet durable sur la confiance : le démarrage avec un faible niveau de confiance peut mener à

une mauvaise utilisation du système, l'opérateur va moins se fier au système et la confiance peut diminuer davantage (EATMP, 2003a).

3. **Minimiser le nombre de petites fautes et d'erreurs dans les tâches faciles.** Les petites fautes qui produisent des résultats imprévisibles (non anticipés) ou les erreurs dans les tâches faciles entraînent un déclin de la confiance plus important que les fautes plus grandes mais permanentes et les erreurs dans les tâches difficiles (Muir and Moray, 1996; Madhavan et al., 2003).
4. **Organiser et formater l'information selon les attentes des utilisateurs envers l'automatisation.** L'engagement des utilisateurs dans le développement du système peut aider à comprendre leurs attentes (Shneiderman, 1998, chap. 2; Robert, 2002).

   Le respect des attentes des utilisateurs renforce le développement de la confiance affective et analogique (Miller, 2005).
5. **Tenir compte des différences individuelles et culturelles dans le design du système et la formation des utilisateurs**.

   Les différences culturelles peuvent influencer la confiance sans avoir de lien direct avec les caractéristiques du système. À cause des différences culturelles, les utilisateurs peuvent avoir des attentes différentes envers le système et ceci peut causer une mauvaise utilisation du système (Zuboff, 1988; Huang et al., 2002, Jiang et al., 2004).

Lignes directrices

1. **Donner accès aux données brutes.** On ne peut pas toujours garantir la fiabilité des systèmes automatisés à cause de l'imprécision des capteurs, des situations anormales, du bruit, etc. (Parasuraman et al., 2000; EATMP, 2003a). Cependant, l'opérateur peut se fier à l'automatisation si les données brutes soient disponibles (Gupta et al., 2001; Wickens & Xu, 2003; Sarter & Woods, 19970). L'accès aux données brutes aidera les opérateurs à diminuer l'effet d'ancrage cognitif (Madhavan & Wiegmenn, 2005). Si les données ne sont pas disponibles, les

opérateurs n'auront d'autres choix que de se fier totalement au système (confiance excessive) ou bien d'ignorer l'automatisation.

2. **Fournir des indications quand les données sont incomplètes, pas fiables, ou incomplètes.**

   Quand les utilisateurs ont conscience des fautes ou des erreurs d'automatisation, ils peuvent développer une stratégie pour compenser ces erreurs et continuer de faire confiance à l'automatisation (Lee and Moray, 1992, Itoh et al., 1999; Dzindolet et al., 1999, 2000a).

   De plus, la confiance est difficile à gagner mais facile à perdre. La méfiance est plus persistante que la confiance. Donc, dans le cas de fautes non anticipées, l'opérateur peut se méfier du système et il faudra beaucoup de temps avant que la confiance au système réapparaisse (Lee & Moray, 1992; EATMP, 2003b).

3. **Rendre explicite l'objectif de l'automatisation**. Au début de l'interaction avec un nouveau système, la confiance dépend d'abord de l'objectif de l'automatisation (Muir & Moray; 1996, Hoc, 2000). Donc il faut montrer explicitement pourquoi l'automatisation a été développée.
4. **Concevoir le système avec une bonne étiquette.** Une bonne étiquette favorise le développement des confiances analogique et affective (Miller, 2005). Inversement, une mauvaise étiquette peut diminuer la confiance et la performance du système (Parasuraman & Miller, 2004; Miller, 2005). Plus le système génère un effet positif, plus l'opérateur va se fier au système. Mais cela doit être fait très soigneusement. Une bonne étiquette ne doit pas compenser une fiabilité insuffisante car cela pourrait entraîner une confiance inappropriée.
5. **Révéler les règles et les algorithmes et utiliser des algorithmes simples.** L'opérateur va se fier davantage au système si les algorithmes sous-jacents à l'automatisation sont compréhensibles (Wiegman, 2002; Wickens, 2003) ou bien si l'opérateur peut retracer la séquence des décisions prises par le système. Il peut aussi être utile de fournir les résultats intermédiaires à l'opérateur (Lee & See, 2004; Sheridan, 1992).

6. **Grouper et isoler les fonctionnalités moins fiables et vulnérables du système.** La méfiance ne se propage pas à des systèmes similaires ou indépendants, mais elle se propage facilement aux autres fonctions du même sous-système (Muir & Moray, 1996).
7. **Si les algorithmes du système dépendent du contexte, montrer explicitement le contexte à l'opérateur.** Les décisions du système peuvent varier dynamiquement lors de changements du réglage du système ou des conditions externes. Donc la perception de ces changements et des relations entre le contexte et les décisions du système va contribuer à ce que l'opérateur maintienne une confiance adéquate.
8. **Montrer la source des fautes dans l'automatisation**. La confiance va moins se dégrader lorsque la source de la faute de l'automatisation est externe à l'automatisation (ex. sabotage, problèmes avec le transformateur) que lorsque cette faute est interne (ex. bogue du logiciel) (Bisantz and Seong, 2001).
9. **Fournir à l'utilisateur une automatisation adaptable**. Une transition des tâches entre l'opérateur et le système gardera l'opérateur dans la boucle. Cela peut contribuer à réduire la sur-confiance, à donner à l'opérateur une meilleure conscience de la situation et à éviter la dégradation de l'habileté de l'opérateur (Miller et al., 2005).
10. **Donner à l'opérateur une formation afin de développer une confiance adéquate.** Cette formation n'inclut pas seulement la familiarisation avec les fonctions et l'objectif de l'automatisation mais elle indique aussi à l'opérateur les fonctions qui sont moins fiables. Il faut former l'opérateur à identifier quand l'automatisation peut être fautive, quand le système peut se dégrader et comment on peut pallier à ces situations. La formation peut aussi aider à empêcher la sur-confiance et à ajuster les différences des attentes dues aux différences individuelles et culturelles (See & Lee, 2004). La formation va favoriser une confiance initiale élevée (Nickerson, Reily, 2004). Par contre, la formation ne doit pas être utilisée pour compenser un design de mauvaise qualité.

11. **Mesurer la confiance dans le système au début de l'utilisation du système mais aussi après l'acquisition d'une certaine expérience avec le système.** L'évaluation subjective basée sur des questionnaires est le moyen le plus commun de mesurer la confiance (EATMP, 2003c, Jian et al. 2000; Madsen and Gregor, 2000; Jiang et al. 2004). En tenant compte de l'importance du facteur affectif dans le développement de la confiance, il est essentiel de mesurer la confiance initiale à l'égard du système. L'échelle de l'évaluation proposée par Jian et al (2000) semble être le plus approprié pour mesurer la confiance initiale. Après avoir acquis un certain niveau d'expérience, le questionnaire de Madson et Gregor semble être plus approprié. Si la langue maternelle de l'utilisateur n'est pas l'anglais, on doit fournir la traduction exacte et les définitions exclusives de chaque terme de l'échelle. Les gens de différentes cultures peuvent comprendre les mêmes termes différemment.

Les résultats expérimentaux relatifs à l'impact de l'utilisabilité de l'interface sur la confiance

La révision de la littérature révèle qu'il y a certains liens entre l'utilisabilité de l'interface et la confiance envers le système.

Cette étude présente des résultats expérimentaux relatifs à l'impact de l'utilisabilité de l'interface sur la confiance. On a formulé l'hypothèse suivante: « L'augmentation de l'utilisabilité de l'interface supporte le développement de la confiance ».

Pour ce faire on a utilisé l'évaluation heuristique et l'évaluation pas à pas l'utilisabilité d'une nouvelle application de la télédétection IDFS (Intelligent Data Fusion System). Pendant l'évaluation on s'est particulièrement attardé aux problématiques qui peuvent avoir une incidence sur la confiance. On a développé des recommandations pour améliorer l'utilisabilité de l'interface, qui ont été implantées par les développeurs de l'application. En particulier, on a amélioré la compatibilité du comportement du système par rapport aux attentes des utilisateurs, le retour de l'information, la clarté visuelle et la

traçabilité. On a concentré notre attention sur l'interface de la tâche 'Définir fonctions de masse'. Premièrement, c'est la tâche la plus importante pour atteindre les buts (selon l'analyse hiérarchique des tâches) et deuxièmement, l'interface de cette tâche avait beaucoup de problèmes de l'utilisabilité (selon les évaluations heuristique et pas à pas) et causait des problèmes aux utilisateurs (selon l'évaluation pas à pas).

Les deux interfaces (initiale et modifiée) ont été testées par six utilisateurs du domaine de la télédétection. Pour mesurer la confiance dans le système, on a utilisé l'échelle de Jian et al. (2000). Pour mesurer la confiance globale dans le système, on a appliqué le modèle proposé dans cette étude. Les résultats ont révélé que, bien que les utilisateurs aient été conscients que les algorithmes et les fonctionnalités des deux interfaces aient était les mêmes, la confiance dans la nouvelle version de l'interface était presque deux fois supérieure (c.-à-d. 1.93) à la version ancienne. Les résultats supportent donc notre hypothèse que l'utilisabilité sert à la confiance.

Les résultats sont aussi cohérents avec la proposition (Miller, 2005) selon laquelle début de l'interaction avec un nouveau système, la confiance dépend fortement des confiances affective et analogique. IDFS est un nouveau système et les utilisateurs n'ont pas d'expérience préliminaire avec ce système. Donc, le changement (ici l'amélioration) de la confiance ne peut pas être basé sur le processus analytique[3], mais surtout sur les processus affectif et analogique.

On conclut que l'augmentation de la qualité de l'interface, en particulier la compatibilité par rapport aux attentes des utilisateurs, le retour de l'information, la clarté visuelle et la traçabilité supportent le développement des confiances affective et analogique.

Les résultats montrent que l'échelle d'évaluation de Jian et al. (2000) est appropriée pour mesurer la confiance initiale à l'égard du system. Il existe une certaine redondance

[3] La confiance analytique est basée sur la confiance qui est développée par analyse du système d'après l'expérience, la connaissance et le modèle mental du système

dans les questions, mais cela permet un meilleur résultat. En calculant la moyenne des réponses, on diminue l'impact des erreurs commises par les utilisateurs telles que celles liées à l'incertitude des réponses ou à la familiarité insuffisante avec le système

Toutefois, l'échelle d'évaluation ne fournit pas de description de la terminologie utilisée. Les tests ont révélé que les utilisateurs ont de la difficulté à comprendre explicitement certains termes (par ex. 'intégrité', 'se fier', 'sécurité'') et à distinguer le sens de certains mots (par ex., la distinction entre 'faire confiance' et 'fiabilité'). On propose, donc de fournir une explication pour chaque item de l'échelle de l'évaluation afin que la perception de la terminologie par les utilisateurs soit cohérente.

On conclut aussi que l'estimation de la confiance globale, suggérée dans cette étude, peut être informative. Pour les deux interfaces, la dispersion ($\sigma \approx 0.077$ pour l'ancienne et $\sigma \approx 0.076$ pour la nouvelle version de l'interface) des valeurs des items différents de l'échelle par rapport à la valeur de la confiance globale est petite. Cela montre que les différents items sont très proches les uns des autres ou qu'ils mesurent des aspects de la confiance que les sujets ont de la difficulté à bien différencier.

L'étude présente aussi les limitations des résultats expérimentaux et du modèle proposé. Il aurait été préférable de mesurer la confiance globale pour un plus grand nombre des tâches et de fonctionnalités du système, et avec un plus grand nombre d'utilisateurs.

Dans le futur, les recherches pourront explorer et mettre en évidence ces différences. Par exemple, une plus grande dispersion peut être un indicateur d'une plus grande confusion par rapport au sens de l'item. Par contre, la dispersion ne peut pas être le seul critère pour assigner des poids. D'autres facteurs devraient être pris en considération, comme l'importance de chaque item pour la confiance.

Une autre piste de recherche serait de mesurer l'impact des autres qualités de l'utilisabilité comme l'aide et la guidage, la cohérence, la prévention des erreurs, la flexibilité et la rapidité de l'utilisation sur la confiance. Une façon de mesurer l'impact

des différentes qualités de l'interface sur la confiance serait de réaliser une expérience avec deux versions de l'interface : l'une avec ces qualités de l'utilisabilité et l'autre sans ces qualités. La comparaison de ses résultats va montrer les qualités de l'interface qui jouent un plus grand rôle dans le développement de la confiance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPEDICES

GLOSSARY

Adaptive system	A system where the flexibility in information or automation behaviour is controlled by the system albeit in service of the human (Opperman, R., 1994).
Adaptable system	The operator is in charge to define when to use the automation and to instruct the system in what behaviours to exhibit (Opperman, R., 1994).
ATC	Air Traffic Control (EATMP, 2003a).
ATM	Air Traffic Management (EATMP, 2003a).
Automation	Device or a system that accomplishes (partially or fully) a function that was previously carried out (partially or fully) by a human operator (EATMP, 2003a).
Automation confidence	Confidence in ability of the machine to support successful completion of the tasks (Taylor et al., 1995).
Automation dependability	The extend to which you can count on the machine to provide appropriate support to the tasks (Taylor et al., 1995).
Automation reliability	The extend to which you can rely on the machine to consistently support the tasks (Taylor et al., 1995).
Calibration	The correspondence between the operator's trust in the system and the capabilities of the automation (Lee and See, 2004).
Cluster	A group of multiband spectral response pattern, which depends on the parameters, the user has differentiated.
Cognitive anchoring	The human tendency to be attached to initially chosen hypotheses among a variety of them (Wickens and Hollands, 2000).
C2	**C**ommand and **C**ontrol
Compatibility	Quality of an interface that respects the expectations and stereotypes of a given population in defined domain (Robert, 1997).
Complacency	A term used to describe the operator's over-reliance on Automation resulting in the failure to detect system faults or errors (EATMP, 2003a).
Consistency	Quality of an interface that respects the rules relative to the meanings, organisation, presentation or behaviour of

	the interface (Robert, 1997).
CPRS	Complacency Potential Rating Scale.
Confidence	Confidence is own ability to successfully complete the tasks with the aid of the automation (Taylor et al., 1995).
Decision aids	Automated systems that provide support to human decision-making processes either unsolicited or by request.
Dependability	The degree to which the behaviour is consistent (Lee and See, 2004).
Dependency	The level to which an operator is willing to depend on a machine (Sheridan, 1988).
DF	Data Fusion.
EATMP	European Air Traffic Management Programme (EATMP, 2003a).
Error management	The functionality of prevention and correction of the errors. The latter includes the detection, the explication and the recovery (Robert, 1997).
Etiquette	Defined roles, acceptable behaviours and interaction moves of human and intelligent agent participants in a common setting (Miller, 2005).
Expert systems	A computer application that uses knowledge base of a human expertise to aid in solving problems.
Explication of intention	Rather than leaving a person in a position of having to understand and discover covert meanings from a system's behavior, the system explicitly displays or says that it will act in a particular way (Sheridan, 1988).
Faith in automation	The extend to which one believes that the machine will be able to intervene and support the tasks in other system states in the future (Taylor et al., 1995).
Familiarity	The feeling of being comfortable with your ability to deal with a situation or object despite a high degree of novelty associated with it (Sheridan,1988). Employment of familiar/friendly/natural procedures, terms, etc. (Llinas, 1998).
Help / guidance	All forms of assistance provided to the user in order to advise, inform, and guide the user during his/her

	interactions with the system (Robert, 1997).
HCT	Human Computer Trust
HDF	Hierarchical Data Format.
HSI	Hyperspectral Imagery.
IDFS	Intelligent Data Fusion System.
Informative feedback	Users should be given clear, informative feedback on where they are in the system, what actions they have taken, whether these actions have been successful and what actions should be taken next (Robert, 1997).
Integrity	The trustee adheres to a set of principles that the trustor finds acceptable. (Lee and See, 2004).
ISO	International Standard Organization.
LM Canada	Lockheed Martin Canada
Mental model	The individual's understanding of the processes underlying system operation.
MSDF	Multi-Sensor Data Fusion
NUREG	Nuclear Regulatory Commission
Opaque system	A system that affords automation users little or no access to raw data on which to base diagnostic decision making (Wiegmann, 2002).
Predictability	The degree to which the future behaviour can be anticipated (Lee and See, 2004).
Pixel	Smallest unit of an image (it is normally square and it represents a certain area on an image).
Reliability	The probability that the system will not fail (Wickens and Xu, 2002).
Robustness	The demonstrated or promised ability of a system to perform in a variety of conditions and circumstances (Sheridan, 1988).
ROI	Region of interest.
SAR	Synthetic Aperture Radar
SATI	SHAPE Automation Trust Index
Self-confidence	Confidence in one's own ability to successfully complete the task (Taylor et al., 1995).

SHAPE	Solution of Human Automation Partnership in European ATM (EATMP, 2003a).
System feedback	All forms of system's replies after the user's action.
Trustor	An individual who establishes trust (in this case, the human operator)
Trustee	One to whom something is entrusted (in this case, the automation)
Understandability	Means that the human supervisor or observer can form a mental model and predict future system behavior (Madson and Gregor, 2000).
Usability	The extend to which the product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use (ISO, 1999).
Usefulness	Classical notion of utility (Llinas et al, 1998). Defines the level at which data or machines respond in a useful way that creates something of value for system users.
Visual clarity	An interface quality, which means that the information on the screen must be clear, well organised, unambiguous and easy to read (Ravden and Johnson,1989).

CHAPTER 1

INTRODUCTION

This chapter briefly presents the current study and the structure of the document.

1.1 Description of the project

The concept of trust is recognized as a subject of increasing importance in the design of complex human-machine systems, in particular in such high-risk domains as military Command Control systems, navy, aviation, process of control of complex automations, etc. The increasing interest in understanding the concept of trust and its impact on safe and efficient utilisation of the automated systems is connected with the increased complexity, sophistication, and also uncertainty of these systems. This memoir emphasizes the importance of the concept of trust, the dangers of undertrust or overtrust that could lead to disuse or misuse of automated systems.

There are two objectives in this memoir:

- The first objective is to propose a set of rules and guidelines for the design of automated systems on how to support an appropriate trust tuning in automated systems. These rules and guidelines are based on an extensive review of theoretical, experimental and empirical results found in the scientific literature. The literature also shows that there are certain links between different properties related to interface usability and trust of operators in the system.

- The second objective is to evaluate the impact of interface usability on the level of trust in the system. The experimental results presented here suggest that an improvement of the interface usability does improve trust. For this purpose a walkthrough and heuristic evaluations of interface usability of a new remote sensing application Intelligent Data Fusion System (IDFS) developed at

Lockheed Martin Canada (LM Canada) have been performed. During these evaluations the emphasis was put on features of interface usability that could presumably affect the operators' trust. The results of such usability evaluations of IDFS have been used to develop recommendations for modifying the IDFS interface. Then these modifications have been implemented by the developers of IDFS. We present the results of measurements of the users' trust (using the measurement scale of Jian et al., 2000) for the old (before modifications) and new (improved) versions of the software. Finally, we discuss the implications of these results and propose improvements of the scale of Jian et al. (2000).

1.2 Structure of work

Chapter 1 defines the goal and the structure of this memoir.

Chapter 2 presents an extensive review of empirical, experimental and theoretical works on trust in automated systems. Here we outline the general concept of trust, its different definitions, and various components of trust. We discuss similarities and differences between trust related to human-human and human-machine interaction, general factors influencing trust, and the cognitive processes underlying the trust tuning. In this chapter we also discuss the dynamics of trust and reliance, the importance of the context and the role of the information displays for the calibration of trust.

In Chapter 3 we discuss the existing tools for the measurement of trust and conduct a comparative analysis. We propose a model for estimating overall trust on the basis of the multi-parameter trust measurement scale of Jian et al. (2000). Based on the literature review in Chapter 2 and on the conclusions derived from the comparative analysis of different trust measurement tools, we propose design rules and guidelines on how to support appropriate trust tuning in new automated systems. We conclude by discussing the relations between interface usability and trust.

Chapter 4 presents an experimental study on the impact of the interface qualities on trust development. Here we present the hypothesis that the enhancement of the interface

usability will supports the development of trust. In this chapter we analyze the software problems of IDFS from a human factors perspective. The steps of IDFS interface usability evaluation are presented. Then the chapter describes the evaluation methods used, and the recommendations we proposed for the enhancement of interface usability. We present the results of trust measurements, using the scale of Jian et al. (2000), for both the old and new (redesigned) versions of the interface. Furthermore, we discuss the relations between usability qualities and trust, and the appropriateness of the Jian et al. (2000) scale for the measurement of trust in novel systems.

Chapter 5 we present the conclusions of this study and propose some research avenues for the future.

In the Appendixes the following supplementary sections are presented:

- Appendix A presents the JDL Data Fusion (DF) model, which was applied in the development of the IDFS application.
- Appendix B outlines different trust measurement scales. In particular, the scales of Madson and Gregor (2000), Singh et al. (1993), Lee and Moray (1994), Taylor et al. (1995) and EATMP (European Air Traffic Management Program) (2003) are presented
- Appendix C presents the scenarios developed for the evaluation of the IDFS.
- Appendix D shows the task hierarchy of the IDFS.

CHAPTER 2

THE CONCEPT OF TRUST IN AUTOMATED SYSTEMS

In this chapter we analyze the importance of appropriate trust tuning in complex automated systems. We present theoretical studies and results of different empirical and experimental studies on trust in automated systems.

2.1. Importance of trust in automated systems

In this information age, information technologies and automation are becoming ubiquitous. Automated systems are increasingly being integrated in complex systems such as C2, aviation, navy, process control, information retrieval, and healthcare. Automated systems provide support to human decision making processes either unsolicited or by request.

The ultimate goal of the new technologies is their safe and efficient utilization by the human operator. One of the major factors that impacts this utilization is the level of trust the operator has in the system (Bisantz et al., 1999; Dzindolet et al., 2002; EATMP, 2003b; Madhavan et al., 2003; Miller et al., 2005; Muir and Moray, 1996). Operators may not use a well-designed reliable automated system if they consider it as untrustworthy (Parasuraman and Riley, 1997).

For example, when in the 80's in two pulp and paper mill plants of American Paper Company a new computer based technology was implemented, the operators did not trust the displays (Zuboff, 1988). They were used to organized teams: "One person staying in the control room, the other roving the plant, and the two communicating via walkie-talkie. Others would run back and forth between the control room and the production area to verify the system's reading". A close collaboration with designers and trainers during four years had helped the operators to trust the systems.

In the mid 80's, the first legal expert systems (rule-based systems and case-based reasoning (CBR) systems) had appeared. Neither of these systems has been developed for a wide-scale use. Both had difficulties in finding fields of application that would lead to a real industrial development. One of the essential factors for those difficulties was the mistrust of operators in these systems (Guidotti and Turchi, 1995).

"A recent survey of Air Force and Air National Guard pilot attitudes regarding the role of UAVs (Unmanned Aerial Vehicle) as wingmen for manned aircraft revealed an inherent distrust in highly autonomous systems. Anecdotal reports from soldiers in Afghanistan using unmanned ground vehicles reveal that soldiers are underutilizing the robots because they inherently distrust the robots" (Cummings et al., in review).

The empirical (Zuboff, 1988) and experimental (Bisantz et al., 1999; Dzindolet et al., 2003; Itoh et al, 1999; Lee and Moray, 1992; Madhavan et al., 2003; Miller, 2005; Muir and Moray, 1996) studies demonstrate that the concept of trust is an important issue in human automation interaction. Their observations and findings show that trust is an attitude towards the automation that affects the reliance on the system. People tend to rely on automation they trust and reject the automation they do not trust. If the system is not trusted, it is unlikely to be used; and if it is not used then an operator may have limited information regarding its capabilities (Muir and Moray, 1996).

Another danger connected with the psychological factor of trust is the operator's overtrust in the system. Overtrust can lead the operator to rely uncritically on automation without recognizing its limitations or fail to monitor the automation's behavior (Parasuraman and Riley, 1997). Because of an excessive trust the pilots failed to intervene and take manual control, which caused the crash of the Airbus 320 (Sparaco, 1995). In automation authors refer to overtrust as "complacency" (Parasuraman et al., 1993; Parasuraman and Riley, 1997; Parasuraman et al., 2000).

The undertrust in the automated systems can result in its disuse, and the overtrust can result in misuse of the system. Misuse and disuse are both inappropriate in automation

and can compromise safety and profitability (Lee and Sanquist, 2000; National Transportation Safety Board, 1997). Hence, to avoid the disuse and misuse of the automated system, it is essential to support the operators to develop an appropriate trust in the system. Appropriate trust means to calibrate mismatches between the level of trust and real capabilities of automation (Lee and Moray, 1994; Muir, 1987).

The interest in trust has grown mostly in the last five years. This interest is connected with the fact that the systems today are becoming more and more complex. The operator may fail to understand the mechanism of the automation, therefore get certain level of uncertainty regardless whether the automation works well. Under such circumstances the operator's level of trust in an automated system is essential (Itoh et al., 1999). Trust helps the people to accommodate the cognitive complexities and uncertainties (Lee and See, 2004).

Although there are many recent studies concerning the role of trust in utilization of the automated systems, some findings are confusing and conflicting (Lee and See, 2004). In certain cases the terminology used is not consistent. There are conceptual models proposed by Dzindolet et al. (2002), Lee and See (2004), and there is a guideline document for a specific domain (for Air Traffic Control systems (ATC), EATMP, 2003a, 2003b), however, there are no established generic guidelines in a larger scope.

2.2 Definitions of automation and trust

In order to understand the concept of trust in automation it is important first to understand what the term automation means.

Automation is a device or system that accomplishes (partially or fully) a function that was previously carried out (partially or fully) by a human operator (EATMP, 2003a).

In this definition the word 'partially' indicates that automation can be applied to different degrees or levels. Hence, an automated system is still a human-machine system.

Trust has several definitions because of diverse interest in it. For human-human interpersonal relationship all definitions for trust include elements of an attitude or expectation regarding behaviours and outcome. Some of these definitions are the following: "expectancy held by an individual that the word, promise or written communication of another can be relied upon" (Rotter, 1967, p. 651) or "expectation related to subjective probability an individual assigns to the occurrence of some set of future events" (Rempel et al., 1985, p. 96)

In other works trust has been defined as willingness to rely on an exchange partner in whom one has confidence or "Willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that party" (Mayer et al., 1995, p. 712).

Many definitions of trust also indicate the importance of the goal-oriented nature of trust. In these definitions trust describes a relationship that depends on the characteristics of the trustee, the trustor, and of the goal related context of the interaction. Analyzing different definitions of trust, Lee and See (2004) draw a conclusion that "trust is the attitude that an agent will help achieve an individual's goal in a situation characterized by uncertainty and vulnerability".

In the context of complex human-machine systems, Madsen and Gregor (2000) have defined trust as follows:

"Trust is the extent to which a user is confident in, and willing to act on the basis of the recommendations, actions, and the decisions of computer-based tool or decision aid."

2.3 General factors influencing trust

In order to provide guidance on how to support an appropriate trust tuning, it is essential to understand the nature of trust. Researches from both social psychology (i.e. human-human interaction) and human-computer systems agree that trust is a multi-dimensional,

dynamic concept reflecting a set of interrelated perceptions and actions of a human. The term multi-dimensional means that trust is a construct that is composed of several different components or elements. Different authors use the term "component" differently, although they all have the same meaning. Some authors call it 'bases of trust' (Lee and Moray, 1992), others call it 'dimensions, factors, components or attributes of trust' (Bisantz et al, 1999; EATMP, 2003a; Jian et al., 1998; Lee and See, 2004), 'constructs or causes of trust' (Madsen and Gregor, 2000). All these terms are factors that influence the development of trust. To avoid the confusion in this study the term "components" will be used.

2.3.1 Components of trust in human-human interaction

According to the model of trust developed by Rempel et al. (1985), in interpersonal relationships trust consists of three components: predictability, dependability and faith. They suggest that trust between individuals has dynamic characteristics and that it would progress in three stages over time: *from predictability to dependability to faith.* Predictability is the degree to which the future behaviour can be anticipated; and dependability is the degree to which the behaviour is consistent (Lee and See, 2004). As the relationship matures, trust shifts to faith. Faith is a more general judgment that the person can be relied upon. Faith describes the aspect of trust or belief that must go beyond the available evidence to accept a given supposition as truth.

2.3.2 Components of trust in human-computer interaction

Zuboff (1988) studied how people trust automated systems in the workplace. She found that people tended to distrust the technology of the automated system and thus used the system less, or that they tended to over-trust the system, which resulted in other problems when the system failed. In the trust building process Zuboff (1988) proposed the following components: *trial-and-error experience*, *understanding of the technology's operation and faith.*

Other empirical studies, consistent with Zuboff's, have shown that people's strategies with regards to the use or non-use of automated aids may be affected by their trust in the

system. Sheridan (1988) examined how trust affects the operator's use or non-use of an automated aid when the opportunity arises. He suggested seven components of trust in command and control systems: *reliability, robustness, familiarity, understandability, explication of intention, usefulness, and dependence.*

- ***Reliability*** implies the reliable, predictable, and consistent functioning of a system.
- ***Robustness*** is the demonstrated or promised ability of a system to perform in a variety of conditions and circumstances.
- ***Familiarity*** is the feeling of being comfortable, somehow familiar with a situation or object despite a high degree of novelty associated with it.
- ***Understandability*** is neither totally the same nor completely different from familiarity. It is one's ability to develop an appropriate mental model of the situation, possibly with the aid of familiarity.
- ***Explication of intention,*** rather than leaving a person in a position of having to understand and discover covert meanings from a system's behavior, the system explicitly displays or says that it will act in a particular way.
- ***Usefulness*** defines the level at which data or machines respond in a useful way that creates something of value for system users.
- ***Dependency*** is the level to which an operator is willing to depend on a machine.

Muir (1987) investigated trust in human-machine interaction, particularly in the operation of supervisory control system. Lee and Moray (1992, 1994) extended Muir's studies. In particular, they showed that operator's reliance and trust in the automation depends on the operators' self confidence in their own abilities.

Muir and Moray (1996) developed Muir's work further and showed that in human-machine interaction for trust development the Rempel et al. (1985)'s model could be applied. They proposed a model with six components: *predictability, dependability, faith, competence, responsibility, and reliability.*

- ***Predictability*** is the degree to which the future behaviour can be anticipated. Reliability is the repeated, consistent functioning of the system.
- ***Faith*** means that the user has faith in the future ability of the system to perform even in situations in which it is untried.
- ***Competence*** means that the system is perceived to perform accurately and correctly based on the input information.
- ***Dependability*** is the consistent behaviour of the system.
- ***Responsibility*** is the expectation that the motives of system designers are reliable.
- ***Reliability*** is the reliable, predictable, and consistent functioning of a system.

Trust can also depend on reputation and gossips, i.e. it can be developed without any direct contact with the system (Burt and Knez, 1996).

Hence, summarising all these studies we can identify the following main elements in trust development:

- Reliability
- Dependability
- Predictability
- Understandability
- Explication of intention
- Competence
- Self-confidence
- Reputation
- Usefulness
- Faith
- Robustness
- Familiarity

Lee and Moray's model for components of trust

Lee and Moray (1992) defined the main factors which influence trust in automation. They identified the performance, process and purpose as the general components of trust.

- **Performance** refers to current and historical operation of the automation, its ability to achieve the operator's goal. Performance information describes *what the automation does.* This is the direct observation of system behaviour.
- **Process** is the degree to which the automation algorithms are appropriate for the situation to achieve the operator's goal. Process information describes *how the automation operates.* This is the understanding of the underlying mechanisms.
- **Purpose** is the degree to which the automation is being used within the realm of designer's intent. Purpose describes *why the automation was developed.* This is the intended use of the system. Purpose corresponds to faith and reflects the perception that the trustee (i.e. the system) has positive intentions toward the trustor (i.e. the operator).

Lee and See (2004) indicate that the performance refers to such trust components as reliability, predictability and robustness; the process refers to dependability, integrity and understandability, and purpose corresponds to faith, i.e. the perception that the trustee has a positive orientation towards the trustor. The operator will tend to trust automation if it performs in a manner which reliably achieves the operator's goals (performance), if its algorithms can be understood and seem capable of achieving the goals in the current situation (process) and if it is designed to achieve the given goals (purpose). Dzindelot et al. (2003) highlight the importance of process. They suggest that the more accurate is the human operators' understanding of how the automated aid operates, the more the operator will trust the automation results. Without a correct understanding of the process the automation uses to attain its decision, human operators may consider an automated aid as untrustworthy. Seong and Bisantz (2000) emphasize the importance of understanding the automated systems inner workings or information regarding how the system generates the environmental estimates. Their experimental results revealed that the informative feedback with information about the inner workings of the automated decision aids increased the operators' understanding of the system and compensated the 'poor' system reliability.

In early relationships with automation, trust can depend on purpose, and not on performance, since there might be little history of performance but clear statement regarding the purpose of automation (Lee and Moray, 1992). This is consistent with the argumentation of Muir and Moray (1996) about the attribute order in developing trust, i.e. trust may first be based on faith or purpose, and only then as experience increases, operators may develop a feeling for the automation dependability and predictability.

Lee and See (2004) suggest that designing interface and training to provide operators with information regarding the purpose and performance of automation could enhance the appropriateness of trust. However, they mention that the availability of this information is not enough to ensure an appropriate trust, and that the information should be presented in a manner which is consistent with cognitive processes underlying the development of trust.

Lee and See (2004) identify that humans develop and tune their trust in three qualitatively different ways: based on analytic processes, analogical processes, and affective, emotional processes. The next section describes the trust development by these three cognitive processes.

2.4 Cognitive processes underlying the trust tuning

Development of trust can be achieved by analytic, analogical, and affective methods and each of them demands a different degree of cognitive processing.

Lee and See (2004) compare an analytic process with the knowledge-based performance, and analogical process with the rule-based performance described by Rasmussen (1983). In analytic process the information is processed and plans are formulated and evaluated using the user's knowledge, experience, and mental model of the system. The analytic method of trust tuning demands the highest cognitive resources of human operator. So far there is no explicit explanation of how trust develops during analytic processes.

Trust developed according analogical judgments, is based on rules and procedures, and it demands less cognitive resources. When rules are consistent with trustworthy behavior, they can increase people's expectation of satisfactory performance by pairing situations with rule-based expectations. This is true during times of normal operations, but rules can have negative effects when situations diverge from normal operations. Therefore, if trust is primarily based on rules, that characterize performance during normal situations, it can be very fragile and might collapse during abnormal situations. Analogical trust can also be based on such factors as reputations and gossips (Miller, 2005). In this case trust will be developed without any direct contact with the system (Burt and Knez, 1996).

The affective aspects of trust represent the core influence of trust on behavior (Kramer, 1999). Emotional aspects are critical, because people not only think about trust; they also feel it (Fine and Holyfield, 1996). Affect may serve as a cue for many important judgments. Readily available affective impression can be easier and more efficient than weighing the pros and cons of various reasons or retrieving relevant information from memory, especially when the required judgment or decision is complex or mental resources are limited (Slovic et al., 2002). Alhakami and Slovic (1994) found that people base their judgments of an activity or a technology not only on what they *think* about it but also on what they *feel* about it. The study shows that if people have positive feeling towards any activity or technology they are inclined to underestimate the risks and overestimate and benefits. And the opposite is also true: if the affective factor is negative, the risks could be overestimated and the benefits underestimated. Nickerson and Reily (2004) propose that the more positive are the affects a machine can generate in a human the more it will be trusted. Emotions can guide behavior when rules fail to apply and when cognitive resources are not available to support a calculated rational choice. However, caution should be taken, since it is possible that the affect may cause the operator to overly trust the machine.

2.4.1 Progress of trust

An important element in trust development is its dynamic character (Lee and See, 2004). There is a temporal element in trust building. It takes time to acquire trust tuning. Trust can be developed via self-experience, training, hearsay or experience of others. Miller (2005) considers that during this process, particularly when interacting with novel system, the affective and analogical processes will be more important in human interaction with automation than the analytic process. When experiencing a novel system for the first time, with no background knowledge about agent's motivations and behavior, the only information the person may have about whether or not to trust, will be the affective information. If the affect is negative enough to prompt a strong "do not trust" response, then, more probably, no further information will be gathered. But when there is a moderate level of affective trust tuning, the user will continue to use the system and develop trust via the analogical process.

Hence, if the system does not provide appropriate cues to achieve at least moderate level of affective and analogical trust, there may never be a chance to build the analytic trust (Miller, 2005). The experimental results (Parasuraman and Miller, 2004; Miller, 2005) also revealed that the analogical and affective methods to tune the trust will play greater role than the analytic method with its more knowledge, and experience. Marsh and Meech (2000) call this affective factor "initial trust inducing stage", and they note the importance of trust building at this initial stage.

2.4.2 Importance of etiquette in trust development

In their review Lee and See (2004) point out the importance of "etiquette" in complex systems, particularly its role in the development of analogical and affective trust. "Etiquette" is the largely unwritten codes that define roles and acceptable and unacceptable behaviors or interaction moves of each participant in common 'social' settings (Miller, 2005). Etiquette rules create an informal contract between participants, and on that basis expectations and interpretations to be formed and used about the behavior of others. As an example, one mechanism for the development of analogical

trust via etiquette is the utilization of a familiar domain terminology. Another example of etiquette is the communication style, which refers to the "interruptiveness" and "impatience" in delivering relevant text messages (Parasuraman and Miller, 2004). Affective influence is connected with the positive and negative reaction produced by the automation etiquette.

The experimental results from Parasuraman and Miller (2004) revealed that both the performance and trust in the system were lowered by poor automation etiquette even when the reliability of the system was high. Studies in psychology have demonstrated that analytic reasoning cannot be effective unless it is guided by emotion and affect, and that the affective process has greater influence on analytic process than the analytic has on affective (Loewenstein et al., 2001; Slovic et al., 2002).

Slovic et al (2002) underline "While we may be able to "do the right thing" without analysis (e.g. dodge a falling object), it is unlikely that we can employ analytic thinking rationally without guidance from affect somewhere along the line". A system that expresses affect will be more trusted by the human, although the human is aware that that machine itself does not experience emotion (Bickmore, 2003; Norman et al., 2003).

The roles of analytic, analogical, and affective processes depend on the evolution of the relationship between the trustor and trustee, the information available to the trustor, and the way the information is displayed.

2.5 Human-human and human-automation trust

In order to understand how to support an appropriate trust development in the system, it is relevant to establish whether human-automation develops in the same manner as the human-human trust.

The early researches concerning human-machine trust were based on the concepts of trust adapted from social psychology, i.e. from studies of human-human interaction.

Some researchers argue that the human-automation interaction is similar to human-human interaction (Bowers et al, 1996). Other researchers suggest that people do enter into relationship with computers, robots and interactive machines in a manner similar to other humans (Nass et al., 1996; Reeves and Nass, 1996). Nevertheless, there are subtle differences in the manner in which people perceive and react to automated aids compared to human teammates.

Muir (1987) constructed a model of human trust in automation by incorporating the components of trust proposed by social psychology. Later research conducted by Muir and Moray (1996) showed certain consistencies between human-human (particularly trust model proposed by Rempel et al., 1985) and human-automation relationship. They show that three components of trust (predictability, dependability and faith) are fundamental characteristics in both human-human and human-machine relationships and that in both cases these components are developed over the time. However, they found differences in the order of the components (i.e. in which order trust develops). According to Rempel et al., in human-human interaction trust starts to develop with predictability, then dependability, and then it turns to faith (See Section 2.3.1.). Muir and Moray argue that in human-machine interaction trust will first be based on faith, then on automation's dependability and predictability. This is consistent with the argumentation of Hoc (2000) that in early relationship with automation, trust can depend on purpose or faith and not performance since there may be little history of performance but clear statement regarding the purpose of automation. And only after having certain experience with the system the operator may develop opinion for automation dependability and predictability.

Lee and See (2004) assert that if "the designer's intent has been communicated to the operator, the latter (the operator) will tend to trust the automation to achieve the goals it was designed to achieve". In interpersonal relationships it can take years of experience to understand the human partner's intention and develop faith in this relationship.

The experimental results of Parasuraman and Miller (2004) confirm that similar to human-human interaction etiquette also impacts the calibration of trust in human-automation interaction.

The findings of Dzindolet et al. (2002, 2003) show that the operators have expectations of "perfection" or high credibility assessments in the case of automation and expectation of "imperfection" or low credibility assessments in the case of humans.

In their conceptual model of sequential trust development in humans vs. automation Madhavan and Wiegmann (2004) show that the machine is perceived as having properties such as 'invariance', or the ability to perform consistently across situations. On the other hand the humans are perceived more adaptable across situations. The interaction of these characteristics with the user's biases will impact the development of trust in the automated system.

Dijkstra et al. (1998) show that students judged advice from an automated expert system to be more rational and objective than the same advice from a human advisor.

The questionnaire study conducted by Jian et al. (2000) show that people tend to be less extreme in their assessments of human-human distrust than trust. The findings show that this is not the case for assessment of human-machine trust and distrust. At the same time Jian et al. (2000) findings show that in spite of these differences people do not perceive the concept of trust differently across the human-human and human-machine relationships.

Analyzing the differences between human-human and human-automation trust, Lee and See (2004) assert that the intentionality of the trustee is an important element of trust between the people. The authors consider that although the automation has the intentionality of the designer, that it is not the same as for the case of human-human interaction.

Another difference that Lee and See (2004) note is that "trust between people is often a social exchange relationship". In human–human interaction a person may act in trustworthy manner to elicit a favourable response from another person (Mayer, 1995). And there is symmetry in the interpersonal trust. How one is perceived by the other influences the behaviour. Lee and See note that "in human-machine interaction such symmetry does not exist".

In their experimental results, Lewandowsky et al. (2000) found that within a complex environment the dynamically changing moment-to-moment trust between people resembles the trust in human-automation interaction. However, it is important to mention that this experiment does not consider the human-human trust in a context of stable long-term relationships. It considers trust during moment-to-moment dynamics between people who share tasks within a complex environment. They found slight differences in task delegation in human-human interaction. People were more likely to delegate the task to human colleagues than to automated systems when they perceive their own trustworthiness to be low. Hence, human collaboration benefits from calibration of people's assessment of how others perceive them.

In their experiment Lerch et al. (1997) provided the students in marketing classes guidance in solving financial management problems. Participants were told that the guidance was an expert system, a human expert or a human novice. The participants expressed not only their degree of agreement with the guidance but also their confidence in it. The results revealed that the participants were more likely to agree with the guidance from an expert system if they were told that the system had already outperformed the human experts. But even then, the participants still expressed greater confidence in the judgments of the human expert. The participants declared that the human can be engaged in something called 'effort' and computers cannot.

Dzindolet et al. (2002) findings show that although self-report data indicate a bias toward automated aids over human aids, performance data revealed that participants

were more likely to disuse automated aids than to disuse human aids. This is consistent with the overall observations that there is a gap between what the users say and what they really do (Robert, 2002).

From these discussions above we can draw the following conclusions:

- In human–automation interaction trust will progress differently than in human-human interaction. It will start with faith or purpose, and then the operator can develop an opinion about the dependability and predictability of the automation.
- Similar to human-human interaction, etiquette does have an impact on the calibration of trust.
- The operators believe that automated systems are more rational and objective. At the same time they have expectations of "perfection" with regard to automation and expectations of "imperfection" with regard to humans.
- The users are not always consistent in what they do and what they say. Although self reports can indicate bias towards the automation, the human operator may more disuse the automation aid than the human aid. And as the opposite, the operators can use the automated expert system but still express greater confidence in the judgments of the human expert.

In the next section we consider the distinctions between trust and reliance, the impact of automation faults on trust. We also discuss how different context parameters interact with trust, and how information display impacts trust.

2.6. Trust and reliance, context, and information display

In developing an appropriate trust three main components should be taken into account: *the dynamics of trust and reliance, the importance of the context, and the role of information display* (Bisantz and Seong, 2001; Dzindolet et al., 2002; Jiang et al., 2004; Lee and Moray, 1994; Lee and See, 2004; Parasuraman et al., 2000; Riley, 1994; Wickens and Xu, 2002).

Trust and its effect on reliance are part of a closed-loop process: on one hand the dynamic interaction with automation influences trust and on the other hand trust influences the interaction with automation. Lee and See (2004) advocate that trust guides but does not completely determine the reliance.

Here it is important to establish the differences between the terms **reliance** and **trust.** Trust is a purely psychological state and can be assessed by subjective rating. Reliance, on the other hand, can be assessed through subjective methods and through performance (Wickens and Xu, 2002). One can rely upon the automation even if it is not trusted, i.e. the user can know that the automation can fail but use it (rely on it) nevertheless, because he/she is overloaded or there is no other possibility to conduct that task. In this case the distrust of automation might increase the workload by forcing the users to become skeptical and increase their monitoring (John et al., 2005).

Trust can be strongly affected by system reliability (Masalonis and Parasuraman, 1999; Moray et al., 2000). Unreliability lowers operator's trust and can therefore undermine the potential system performance benefits of the automation (Parasuraman and Riley, 1997). A highly reliable automation provides the operator the opportunity to observe how the automation works and thus to develop trust in the system.

Based on different theoretical and experimental results Lee and See (2004) analyze the relations between trust and different automation levels. Parasuraman, Sheridan and Wickens (2000) assert that *trust depends on the type of automation.* The automation can differ in type and complexity, from simply organizing the information sources, to integrating them in some summary fashion, to suggesting decision options, or even to carry out the necessary action. In their taxonomy Parasuraman et al. (2000) provide four levels (stages) of automation, which can vary depending on situational demands during operational use:

- Information acquisition: collection and filtering of information.

- Information analysis: integration of information in a form of inference or diagnoses.
- Decision and action selection: selection among decision alternatives.
- Action implementation: execution of chosen action.

The authors mention that high reliability cannot be always guaranteed because of the uncertain nature of information sources, sensor imprecision or conditions in which the algorithms used by automation are inappropriate. Nevertheless, it is possible for the operator to observe the behavior of information acquisition, i.e. stage 1, even if they are not relying on it (Yeh and Wickens, 2001). At this stage the raw data are still available, so here the operators can balance their attention between the raw data and automation (Wickens et al., 2000). In contrast, it is not possible for the operators to observe action implementation, i.e. at stage 4, unless they are relying on it (Lee and Moray, 1994).

Low perceived reliability of the system, especially in opaque complex systems (i.e. a system that affords automation users little or no access to raw data on which to base diagnostic decision making) might increase the operator's cognitive anchoring effect. Cognitive anchoring is the human tendency to be attached to initially chosen hypotheses among a variety of them. In opaque systems the operators may guess the cause of the system states and anchor their agreement with system decision according to this initial "guess". The cognitive anchoring effect will increase the probability that the operator will disagree with the system results, resulting in its distrust and disuse (Madhavan and Wiegmann, 2005).

2.6.1 Impact of automation faults on trust

Trust and reliability are strongly affected by faults in the automation (Dzindolet et al., 2002; Jiang et al., 2004; Lewandowsky et al., 2000; Tan & Lewandowsky, 1996). As mentioned in Section 2.5, humans expect that the automated systems should perform in near perfect rate, whereas in the case of humans they have expectations of "imperfection" (Dzindolet et al., 2002; 2003). The researchers in cognitive psychology

have found that the information inconsistent with expectations (e.g. schemas) is likely to be well remembered and play an unjustified role in information processing (Ashcraft, 1994). Therefore, the system errors will be inconsistent with human expectations towards the automation, and these errors are more likely to be remembered than the correct system decisions. All this may distort the operator's estimate of the system's reliability and trigger a rapid decline in trust when automated system makes errors (Dzindolet et al., 2002; 2003; Madhavan and Wiegman, 2003).

In this regard Riley (1994) underlines the importance of prior knowledge of the fault. People rely on faulty automation when they are aware of automation imperfection. When the operators are aware of the automation faults or imperfection, they can more carefully calibrate their allocation of attention (automation stages 1 and 2) or preparation for incorrect actions (automation stages 3 and 4) to that imperfection (Wickens and Xu, 2002). The importance of feedback to an operator about automation errors was also shown in studies conducted by Simpson (1992, 1995) in the domain of naval Command and Control systems. The same findings have been shown by Dzindolet et al. (1999, 2000, 2003). They found that providing operators with information about the conditions in which an automated aid is likely to make errors leads to improved task performance.

St. John and Manes (2002) assert that unreliable automation can be useful if the user interface will respect the principle "trust but verify", i.e. the interface conveys:

- Information about when and how the automation may become less reliable or uncertain.
- Information about what aspects of the automation are more reliable or certain than others.

So if designers want to increase the reliance and trust in automated systems, they should ensure that operators understand when and why the system is likely to make an error.

However, Dzindolet et al. (2003) findings show that simply providing reasons why the aid might make errors is not the fail-safe (the best) solution. The danger is that trust and

reliance may be increased to an inappropriate level, resulting in distrust (and misuse) of the aid. They propose that comprehensive instruction on the algorithms used by the automation may be effective in explaining not only why the automation might make an error, but also how it arrives to correct decisions. The importance of the human operator's understanding of how the automated system operates is mentioned in the conceptual trust model provided by Lee and Moray (1992). As we can see in Section 2.3.2, one of the bases of trust development is the process: operator need a correct understanding of the process a decision aid uses to attain its decision.

In their experimental studies Muir and Moray (1996) concluded that trust is not a discreet variable and that the variable level of trust can vary between none and total. In this variation the trust calibration can be supported by the explanation when the system would and when it would not be correct.

Another interesting finding concerning reliance and automation is that when a systematic fault occurs, the operator can develop a control strategy (Lee and Moray, 1992; Moray et al., 2000). Thus a small fault with unpredictable results affects trust more than does a large fault of constant error (Muir and Moray, 1996). Similarly, an inconsistency between the operator's expectations and the behavior of the automation can undermine trust even when automation performs well (Rasmussen et al., 1994).

The experimental results of another study (Madhavan et al., 2003) revealed that the users' interaction with an aid that makes errors on easy tasks results in a greater reduction in trust and reliance, than when interacting with an aid that makes errors only on difficult tasks while reliably performing easy tasks.

Trust and reliance are dynamic; they will change depending on the user's experience with the system. Initial experience has a lasting effect on trust: an initially low level of reliability leads to lower trust. Trust is more resilient if automation reliability starts high (Fox and Boehm-Davis, 1998). The result of a "low start of trust" can be that the distrusted system is not used and so becomes even less trusted and less used.

Another interesting question is whether distrust in a particular part of a system will spread to the whole system. Muir and Moray (1996) showed that distrust in a particular function might spread to other functions performed by the same subsystem. This may lead to unwarranted distrust, unnecessary monitoring and overriding of good decisions. However, distrust did not spread across independent but similar systems.

2.6.2 Impact of context on trust

As mentioned above, *trust depends on the context*. It is important to understand how the different context parameters interact with the trust. These context parameters are: workload, multitask demands, time constraints, varying levels of risk for the consequences, familiar or unfamiliar situations, self-confidence level, stress, individual differences, stereotypes, training, etc. These trust-context interactions can affect the reliance.

A particularly important variable that interacts with trust and influence reliance is *self-confidence.* When operators' self-confidence is higher than their trust in the system, they are more inclined to rely on manual control. The opposite is also true: low self-confidence is related to a greater inclination to rely on the automatic controller (Lee and Moray, 1994). Thus, the biases coming from the operator self-confidence can have a substantial effect on the appropriate reliance on automation.

Muir (1987) found that people trust the automation more during *unfamiliar situations*. In 1994 he also showed that the users weight each case differently, depending on the *risk involved.* The risk-level and size of consequences of an automation error impact the users' trust (Masalonis et al., 1998, Jiang et al., 2004). Thus, relying on an automated system with a high level of risk consequences calls for a large amount of trust. Besides, Riley (1994) found that it takes longer to recover after automation failure in high-risk situation than in low-risk situation.

The *multitasking demands or high workload of a situation* can also interact with trust to influence reliance. A situation when the operator uses a highly reliable (however, not

perfectly reliable) automation, combined with the responsibility for multiple tasks in addition to monitoring the automation, can lead to an over-trust in automation. This will undermine the detection of automation failures by the operator (Parasuraman et al., 1993). Experiments carried out by Mosier et al. (1998) also show that in multitask environment the highly reliable system can reinforce the perception that the task can be completely carried out by the automation and that there is no need to pay attention on other cues. They call this phenomenon automation bias. Ultimately, the over-reliance or excessive trust (complacency) leads to less vigilant monitoring.

In order to reduce the over-trust in the automation and increase detection of failures Parasuraman et al. (1996) propose the utilization of adaptive automation, i.e. shifts between manual and automatic control according to the capabilities of the person and the situation. Here it is important to mention that some authors distinguish between adaptable and adaptive systems. In adaptive automation the flexibility in the automation behavior is controlled by the system. In adaptable automation the operator can define when to use automation and to instruct the system for what behaviors to exhibit (Opperman, 1994). The advantages and disadvantages of these two automation types regarding the trust, complacency, situation awareness, skill degradation, and performance, are provided by Miller et al. (2005) in their implications. They concluded that requiring operators to make decisions about when to use automation and to instruct the automation in what behaviors to exhibit should produce better trust tuning, better decisions about the automation reliance. Besides, by keeping the operator active in charge of how much and what kind of automation to use and when to use it, we keep the operator "in the loop". This will help to avoid the degradation of skills or situation awareness of operators, which is critical in the case that system failure will take place and the human intervention is needed.

2.6.3 Impact of environmental context on trust

The *environmental context* also influences trust and reliance. Automation may perform well in certain circumstances and not in others. For this reason, appropriateness of trust

often depends on how sensitive people are to the influence of the environment on the performance of the automation. In this regard Bisantz and Seong (2001) found that trust could be affected differently depending on the *source of automation failure*. Their findings show that trust is less degraded if the source of failure is an abnormality outside of automation system itself (different environmental situations as a lost power supply, intentional sabotage versus hardware in military domain), than if it is a failure within the automation (software bugs).

It is also important to mention the individual, organizational and cultural context (Lee and See, 2004). Individuals who are "trustworthy" tend to have more trust in others and are more willing to delegate tasks to automated systems. The experimental results of Jiang et al. (2004) reveal the importance of consideration of individual differences. As mentioned above the individual differences can differently influence trust in the situation of stress. The organizational context reflects the interaction between people. This includes reputation and gossips that will affect the individual's willingness to trust. The cultural context influences trust through social norms and expectations. For example, the level of social trust varies substantially among countries. This variation accounts for over 64 % of the variance in the level of Internet adoption (Huang et al., 2002).

Above mentioned context related issues of trust can be summarized as follows: displaying the context to the operator can lead to better trust development.

2.6.4 Impact of content and the format of the display on trust

Another important factor in trust development is the *impact of the content and the format of the display*.

Perception of the automation-related information is usually mediated by the display. Organizing information on the display in a way that supports analytic-, analogical-, and affect-based assimilation of this information may be an important means of guiding appropriate expectations regarding the automation. If the information is not available in the display or if it is formatted improperly, trust may not develop appropriately.

However, the amount of the information must be tailored to the available decision time (Entin et al., 1996).

In many cases, *trust and credibility depend on surface features of the interface* that have no obvious link to the true capabilities of the system (Briggs et al., 1998; Tseng and Fogg, 1999). For example, Kim and Moon (1998) found that certain interface features such as cool pastel colors and balanced layout could enhance users' perception of interface trustworthiness.

As it is mentioned in the section 2.4, the affective and analogical processes are very important in early stages of trust development. Both these processes are strongly connected with the user expectations with regard to the system. So it is vital to organize the human-computer interface (HCI) in a way that is consistent with user expectations. Of course, the vast majority of software developers intend to provide well-organized and usable systems. But one should note that the viewpoint of the developer could be different from the user's one (Shneiderman, 1998, chap.2). In order to develop a good and useful human-computer interface, it is essential to know the user. In the different methodologies of user-centered design (e.g., Gould et al., 1997; Kretzberg, 1996; Mayhew, 1999; Nielsen, 1993; etc), and in the repertoires of ergonomics guidelines and principles for system design(Smith and Mosier, 1986; MIL-STD-1472; NUREG-0700, 1996), the main message always is “Know the user”, i.e. know the user's profile, tasks, goals, and needs (Robert, 2002).

Hence, to find out the answers to the questions "what information?”, “when?” and “how to present it?” one should involve representative users in the design process in order to validate the HCI with them.

CHAPTER 3

MEASUREMENT OF TRUST, DESIGN RULES AND GUIDELINES

In this chapter we compare the existing tools for measuring trust. We propose a model for the estimation of an overall trust on the basis of the multi-parameter trust measurement scale of Jian et al. (2000). Based on the literature review and results of comparison of the trust measurement tools, we propose design rules and guidelines to support appropriate trust tuning in new automated systems. We conclude by discussing the relations between interface usability and trust.

3.1 Measurement of trust

Trust can be considered as an 'enabler' to the introduction of new systems. Therefore, it is useful to measure operators' trust in the system.

The level of trust in the system can be measured by objective or subjective methods. A very simple objective measure is the frequency of use of a certain automated tool, certain functionality, assuming that if the tool is not used, then it is not trusted. However, it can happen that the system or the automated tool is trusted, but the user does not use it for other reasons than trust. It might be assumed also that if the tool is used or activated, so the user trusts the tool. But it can happen that the tool is activated or used but the user does not trust that information. In psychological terminology trust is an *intervening* variable because it 'intervenes' between particular stimulus conditions and particular behaviours. That is, it is an internal state that cannot be measured directly (EATMP, 2003a). Wickens and Xu (2002) consider that trust is a purely psychological state and can be assessed only by subjective rating. Evidence of empirical studies also indicates that the utilisation of subjective questionnaire-based rating scales is the most common methods to measure trust (Madson and Gregor, 2000, Jian et al., 2000, EATMP, 2003c).

Hence, the subjective measurement of trust appears to be a more appropriate approach. It will consist in asking questions to operators.

Rampel et al. (1985) have been among the first to design a questionnaire to measure trust for human-human interaction.

For the case of human-machine interaction, Lee and Moray (1994) have suggested a 10-point rating scale for 6 items to evaluate the operators' overall trust (single-rating) (see Appendix B.3). The score varies from 1 ("not at all") to 10 ("completely") and is a clear analogue to the NASA TLX workload measures for Air Traffic Control (ATC). The scale items either refer to an entire system or to parts of the system.

Singh et al. (1993) developed a rating scale to measure the humans' potential for complacency in a variety of automated systems by investigating attitudes toward everyday automated devices such as automated teller machines (see Appendix B.2). The scale is called Complacency Potential Rating Scale (CPRS).

Taylor et al. (1995) suggested a 7-point rating scale questionnaire with 17 items (see Appendix B.4). The scale is part of extensive studies on the "human-electronic crew" in the military domain. The scale questionnaire was developed to determine operators' views on the timeliness and appropriateness of adaptive computer aiding.

Muir and Moray (1996) proposed multiple rating scales to extract dimensions of trust. They concluded that trust is not a discrete variable, but that the trust level can vary from 'none' to 'total'. The operators were asked specific questions with regard to experimental system, particularly to rate their degree of trust in three aspects of the system and also to rate the system's performance according to six components of trust: competence, predictability, dependability, responsibility, reliability over time and faith in future ability.

All the questionnaires, mentioned above were created by using adaptations of interpersonal measurement tools or were created by researchers themselves. They have not been empirically derived or validated.

Madsen and Gregor (2000) developed a scale for measuring trust in computers, called Human-Computer Trust (HCT) scale (see Appendix B.1). In this study, a group of subjects identified components (namely 10 trust components) of trust that they believed would affect their level of trust in intelligent decision aid system. Following refinement and modification, the number of components was reduced to five main components: reliability, technical competence, understandability, faith, and personal attachment, each with five sub-items. The authors claim that the HCT has been empirically shown to be valid and reliable. However, the scale has the following limitations (Madsen and Gregor, 2000):

- The results of this study may not be generalized too widely, since the study is context specific. HTC deals specifically with intelligent systems, which are designed to help decision-making. The scale is created with regard to the systems that either provide advice, or make decisions subject to the user's discretion.
- It is also assumed that the users' experiences with the system being investigated may range from a few months to several years. For example HCT contains items such as "I would feel sense of loss if the system was unavailable and I could no longer use it", "The system responds the same way under the same conditions at different times" or "The system uses appropriate methods to reach decisions" (see Appendix B.1).

Jian et al. (2000) have provided the first empirical evidence that trust and distrust can be measured using the same rating scale. The scale was developed as part of a three-phased experimental study. In the first phase, a word elicitation study, various words (38 words) related to the concepts of trust and distrust were collected by seven linguists from questionnaires used in previous studies (from 138 trust related words). In the second phase, Jian et al. (2000) investigated how closely each of these words was related to trust or distrust. This investigation was based on ratings by 120 participants.The third phase was a paired comparison study, in which 30 participants rated the similarity of pairs of words (a total of 435 comparisons). Data from both the questionnaire study and the

paired comparison study were then used to construct a multi-dimensional measurement scale for trust. The result of these three phases is a 12-item questionnaire. The scale incorporates a seven point rating scale labelled from "not at all" to "extremely" (Figure 4-3). The scale proposes generic questions, and there are no specific requirements related to the experience with the system.

Based on the review of the rating scales described above, a new rating scale, called SATI (SHAPE Automation Trust Index), was proposed by the Air Traffic Management (ATM) Human Resources Unit of EUROCONTROL for measuring the controllers' trust in ATM systems during real-time simulations. This scale has been developed by SHAPE (Solutions for Human-Automation Partnerships in European ATM) Project. However, this rating scale is context specific. It is developed particularly for ATC (Air Traffic Control) computer-assistance tools.

The evaluation of SATI in real-time simulation revealed that care is needed to employ exact words in the questionnaire. The words such as 'reliability', 'accuracy', and even 'trust' might mean different things to different nations. For example, the experiments revealed that the controllers' view of trust and confidence was not the same as in the scientific literature. What the authors of scientific literature consider to be 'trust', the controllers consider it as being 'confidence'.

3.1.1 Comparative analyses of different measurement scales

As mentioned in the Section 2.5, trust is dynamic. Hence, it can vary when measured at different stages of use of the same system. Considering the importance of the affective trust in case of new systems, it is essential to measure the initial trust in the system. Here arises this question: "Which scale is more appropriate for initial trust measurement?" The considerations for such measurement are the following:

- At the initial stage the operator cannot have enough experience about the performance of the system. Hence the scale should contain more generic questions.
- Because of the lack of experience at the initial stage of use of the new system the operator may have contradicting feelings towards the system. At this stage the operator's judgement would be based on the affective and analogical processes rather than on the analytical process. Hence, asking the operator only a single question to evaluate his/her trust in the system could not produce a reliable answer. Therefore, it would be preferable to have a multiple rating scale. Multiple rating scale allows one to increase the reliability of the measurement of the trust parameter and to decrease the uncertainties connected with the contradictory feeling of the user with respect to the novel technologies.
- It is preferable to have a scale based on experimental studies.
- It is preferable to have a generic scale that is not restricted to a specific domain.

Table 3.1 shows the existing scales for trust measurement in complex systems and their conformity our selected criteria.

Table 3.1: Comparison of different trust measurement scales

Scales	Is based on empirical results	Is a multiple rating scale	Permits to measure initial trust	Was developed for a specific domain
Lee & Moray (1994)	No	Yes	Yes	Yes
Muir & Moray (1996)	No	Yes	Yes	Yes
Taylor (1995)	No	Yes	No	Yes
Madson & Gregor (2000)	Yes	Yes	No	No
Jian and al. (2000)	Yes	Yes	Yes	No
SATI (2003)	Yes	Yes	Yes	Yes

Taking into account the considerations above, we can draw a conclusion that for the initial stage of utilization of the system the scale developed by Jian et al. (2000) seems more appropriate. This is a multiple rating scale, based on empirical results. Its items in the scale are rather generic, hence the scale can be completed after only a few initial trials of the system. Besides it contains both positively and negatively trust-related items, which could be useful to measure more correctly the affective trust in the system.

However, there is one essential point to mention when using the scale of Jian et al. (2000) scale. As a result it provides 12 answers for 12 different trust-related components, as far as Jian et al. (2000) consider trust as a multi-variant complex concept. Although that seems a correct general proposition, it does not help to answer unambiguously the item "Is the system trustworthy or not?" On the basis of this we can estimate the level of trust in the system. Therefore, for practical reasons it seems necessary to be able to unify somehow the results of all 12 answers, so that one could have an estimate for an overall final trust relative to the system. In the next section a model for this unification is provided.

After reaching a certain level of experience with the system, the HCT scale could be more applicable. It contains more detailed questions about the system reliance. At this stage the operator could have enough experience to answer these questions.

3.2 Measurement of an overall trust with the multi-parameter trust measurement scale.

As mentioned in the previous section, among the currently existing trust measuring scales the multi-parameter scale developed by Jian et al. (2000) (Figure 4-3) seems to be the most appropriate at the initial stage of utilization of a new system. One of the reasons is that the presence of some redundancy and of opposite (positive and negative questions related to trust) trust-related items in this scale would help to reduce the impact of the mixed feelings that an operator would have in case of interaction with a novel automated system. At the same time, we believe that it would be helpful to be able to estimate, on

the basis of all the answers obtained, the level of overall trust of the operator in the system. Redundancy in the items ('questions') would then also help to calculate more accurately both this overall (i.e. the average) trust level and the uncertainty (the 'dispersion') in the received answers, which will be possible to do even for the single user.

The Jian et al's. scale (2000) contains 12 trust related items incorporating a seven-point rating scale where possible answers are integers in the range from '1' to '7'. In this scale '1' is equivalent to 'not at all' and 7 means 'extremely'. The measurement scale contains five negatively and seven positively trust-related questions. The empirical study of Jian et al. (2000) confirms that trust and distrust are oppositely related and could be measured on the same rating scale. This notion suggests that for calculating the overall trust it is reasonable to use a weight system where different items would be ascribed different weights varying from -1 to +1. In principle, the question of assigning the weight for each trust-related item needs a separate study for different automated systems and significant statistics, which is out of the scope of this work.

In this work as a first step for the development of the unification scheme we suggest to unify the positive and negative trust related items. We can give a positive value '+1' to an item bearing positive meaning for trust, and a negative values is '-1' to a negatively trust-related item.

Here we also consider the possibility that some questions from the questionnaire could be simply inappropriate for a specific domain or specific cases. It can also happen that the operator finds the question ambiguous. Hence it seems reasonable to discard answers to such ambiguous or poorly understood questions, since these answers would be more misleading than informative.

Here are the steps for the estimation:

1. Discard the answers for items that appeared confusing or ambiguous for the operator, that is the items with the assigned '0' weight. The number of such items will be $K \leq 12$.

2. Using the answers obtained from N users, calculate the mean values $\overline{a_i}$ for each of these K items:

$$\overline{a_i} = \frac{1}{N}\sum_{i=1}^{N} a_i \qquad \text{i} = 1, 2, \ldots \text{N}.$$

3. Each of the items $\overline{a_i}$ will be then in the range from 1 to 7. It is reasonable to normalize the seven-point rating scale to the interval from 0 to 1, such that the answers to individual items and the overall trust could be measured in percents. The answer "not at all" should then correspond to value 0, whereas the answer "extremely" effectively means 1 (i.e. 100%). Hence, the normalized values can be calculated by the formula:

$$P_i = \frac{\overline{a_i} - 1}{6} .$$

Since $1 \leq \overline{a_i} \leq 7$, we will have $0 \leq P_i \leq 1$.

4. Now we need to correlate the negative and positive trust-related items using the system of positive and negative weights. It is worth noting that there are 5 negatively trust-related items in the Jian et al. (2000) measurement scale that should be ascribed the weight -1.

For a negatively trust related item the value P_i =0 effectively means "full trust", whereas P_i =1 would imply absolute absence of trust. Therefore measuring this parameter in the same "positive trust" scale implies

$$I_i = 1 - P_i ,$$

whereas for a item with weight "+1" we should simply write $I_i = P_i$.

5. After renormalizing and bringing all K items into a single trust scale, the average (overall) trust can be calculated as:

$$T = \frac{1}{K}\sum_{i=1}^{K} I_i$$

6. We can also calculate the variance of this value:

$$\sigma^2 = \frac{1}{K-1}\sum_{i=1}^{K}(I_i - T)^2$$

The dispersion σ (the square root of the variance) will allow us to conclude about the range of variations, of the derived overall trust value.

Note that this procedure will allow to measure the overall trust even for a single user, i.e. when N=1.

3.3 Design rules and guidelines

Based on the review of existing empirical and experimental results and of theoretical works presented in the previous sections (Sections 2.1 to 3.1), design rules and guidelines on how to support appropriate trust tuning in new automated systems are proposed.

The design rules which are typically in line with good software development practices should be helpful for designing decision aid systems. The design rules represent the basic and more important lines, principles based on the author's judgment. The "design rules" can be described as principle lines that are not necessarily applicable; however, they show the direction for more explicit guidance, i.e. the guidelines.

General design rules

1. **Design the system for an appropriate level of trust, that is neither too high nor too low.**

   Neither distrust (caused by an undertrust) nor complacency (resulted by an overtrust) is desirable. It can compromise safety and profitability (Lee and Sanquist, 2000; National Transportation Safety board, 1997; Parasuraman and Riley, 1997; Parasuraman and Miller, 2004; Zuboff, 1988).

2. **During initial introductory periods with new systems prepare both the system (hardware and software) and the users for eventuality of some form of system failure. However, remember that initial experience has lasting effect on trust.**

Despite the pre-operational testing of novel systems the failure rate in these cases is much higher. However, even when the reason is known and understood (e.g. unstable platform, software bugs, etc.) it is difficult to trust a system that keeps breaking down. The "low start of trust" can lead to the situation where the distrusted system is not used. Therefore it becomes even less trusted and less used (Fox and Boehm-Davis, 1998). Hence, even though preparing both the users and system for possible system failures, by the time the formal training period starts the system needs to be virtually 'bug-free' (EATMP, 2003a).

3. **Minimise the amount of small faults and errors in easy tasks.**

   Errors in easy tasks result in a greater reduction in trust and reliance than errors in difficult tasks while reliably performing easy tasks (Madhavan et al., 2003). Small faults with unpredictable results and affect trust more than a big fault of constant error does (Muir and Moray, 1996).

4. **Organize and format the information according to user expectations towards automation.**

   The involvement of real users in system the development could help understand the users' expectations (Robert, 2002; Shneiderman, 1998, chap.2).

   The HCI consistency with user expectations will support the trust development via affective and analogical processes (Miller, 2005).

5. **Consider individual and cultural differences in both designing the system and in training the operators.**

   Individual and cultural differences may influence the reliance in ways that are not directly related to the characteristics of the system (Huang et al., 2002; Jiang et al., 2004; Zuboff, 1988). Cultural differences regarding the expectations towards the system can be a powerful force that can lead to misuse or disuse unless addressed by appropriate training (See and Lee, 2004).

Guidelines

1. **Provide access to raw data.**

   High reliability of automation cannot always be guaranteed, since there would always exist conditions under which the algorithms could be inappropriate, or because of inadequate sensor precision (Parasuraman et al, 2000). Nevertheless, the operator can still rely on automation as long as the raw data are available (Gupta et al., 2001; Sarter and Woods, 1997; Wickens and Xu, 2003). Access to raw data can also help to reduce the cognitive anchoring effect, and consequently it will support the operator to avoid unjustified disagreement with the system results and develop distrust in the system (Madhavan and Wiegmenn, 2005). The danger of unavailability of raw data is that the operators have no other choice but to either totally rely (misuse) on automation or ignore it (disuse).

2. **Provide means to indicate to the user that data are missing, incomplete, unreliable, or invalid.**

   Initial expectations of perfect automation performance can lead to distrust and disuse when the operator finds the automation to be imperfect (Dzindolet et al., 2002, Madhavan et al., 2003; Madhavan and Wiegmann, 2004). Since there will always be conditions where the automated systems can fail (Parasuraman et al., 2000; EATMP, 2003a), the user needs to be aware of the problems, to understand why, and under what conditions an automation tool can make an error. When the operator is aware of the automation error, he/she can develop a control strategy and compensate for such errors (Dzindolet et al., 1999, 2000; Itoh et al., 1999; Lee and Moray, 1992). Hence, the users will trust even this faulty automation (Dzindolet et al., 2003; St. John and Manes, 2002; Relay, 1994).

   Distrust is more resistant to change than trust. Trust is difficult to achieve and maintain, yet easy to lose, and difficult to recover once lost (EATMP, 2003b; Lee and Moray, 1992;). Therefore, once the system is distrusted it will take long time to regain trust in the system.

3. **Make clear to the user the purpose of the automation, the intentions of designers in developing the system.**

   In early interaction with automation, trust largely depends in the first place on the purpose of the automation (Hoc, 2000; Muir and Moray; 1996). Make it clear why the automation was developed. Design the application in such a way that the provided functionalities are related to the user's goals.

4. **Design with good computer etiquette.**

   Etiquette is a promising line to tune trust through affective method (Miller, 2005). The more positive effect a machine could have on the operator, the more it would be trusted. A reliable system with poor automation etiquette can lower the user performance and trust, and good automation etiquette can compensate for lower automation reliability (Miller, 2005; Parasuraman and Miller, 2004). Intentionally developing good computer etiquette can promote trust in the system. However, it should be done carefully. Insufficient reliability should not be compensated by good automation etiquette. This could lead to an inappropriate trust.

   Examples of computer etiquette are the manner the feedback or the information is provided to the user (how the computer should interrupt the user with information), the medium of human-machine communication (e.g. how the message is delivered), or the utilisation of familiar terminology, positive and polite computer feedback.

5. **Reveal the rules and algorithms used by the automation, and keep the algorithms simple.**

   The automation will be more trusted if operators understand the algorithms underlying the automation (Wickens, 2003; Wiegman, 2002). Such understanding provides greater sympathy for "why automation fails". Provide a possibility for the operator to track the sequence of decisions made by the system. It will also be helpful to show the process and algorithms of the automation by providing a tool to display intermediate results in a way that is comprehensible for the operators (Lee

and See, 2004; Sheridan, 1992). Informative feedback on how the automation generates the environmental estimates can compensate the 'imperfect' automation reliability (Seong and Bisantz, 2000). The comprehensive instruction on the algorithms, in particular how it arrives to correct decisions, will support the operator to avoid the over-reliance towards the system as well (Dzindolet, 2003).

6. **Group and isolate less reliable or vulnerable functionalities of your system if it is possible.**

   Separate such functionalities from the reliable ones and make the operator aware of these different functionalities. Distrust does not spread across similar but functionally independent systems. However, distrust in a particular function might spread out to other functions performed by the same subsystem (Muir and Moray, 1996).

7. **If algorithms of the system are context dependent, make the context explicit to the operator.**

   For example, decisions made by the system may vary dynamically when changes occur in system settings or external conditions used by the automated processes. Awareness of those changes – and of the relationship between the context and the system decisions – will help the operator maintain a higher level of trust.

   It would be helpful to provide the operator withinformation concerning time constraints, level of risk, environmental constraints, etc.

8. **Show the source of automation failure.**

   If possible, reveal the environmental situation in the case of automation failure: whether the source of the failure is an abnormality outside of automation system (different environmental situations as a lost power supply, intentional sabotage versus hardware in military domain), or within the software (software bugs) (Bisantz and Seong, 2001).

   Trust would be less degraded if the source of the failure is outside of automation

(e.g., lost of power supply, enemy sabotage) rather than in the automation itself (e.g., software bugs).

9. **Provide the user with adaptable automation.**

   By keeping the operator in active charge of how to use the automation, we keep the operator "in the loop", hence more aware of how the system is or should be performing (Parasuraman et al., 1996). This will help to reduce the complacency and also will support to avoid the skill degradation (Miller et al., 2005).

10. **Train the operator in order to develop an adequate trust.**

    The primary goal of the training should not be "make the operator trust the system" but "to develop trust at an appropriate level" in order to avoid the consequences of operator's over-reliance or under-reliance on automation (EATMP, 2003a). Training should include not only the familiarization of the operator with the provided functionalities and the purpose of the automation, but also revealing to the operator less reliable functionalities. This should also include training for recognition of where the automation could be unreliable, for possible system degradations and means/ways for recovery in these cases. Training can help to prevent the automation bias, hence the complacency. It should also address differences in expectations related to individual and cultural differences (Lee and See, 2004). Training will support the development of higher initial trust (Nickerson and Reily, 2004). However, training should not be used to compensate for a poor design of decision aid system.

11. **Measure trust in the system both at the introductory stage and after acquiring a certain level of experience with the system operation.**

    Subjective questionnaire-based rating scales are the most common means for measuring trust (EATMP, 2003c, Jian et al., 2000; Jiang et al., 2004; Madsen and Gregor, 2000). Taking into account the importance of the affective trust in case of new systems, it is essential to measure the initial trust in the system. The scale

developed by Jian et al. (2000) is more appropriate to use at the initial stage of utilization. After reaching a certain level of experience with the system, the HCT scale (Madsen and Gregor, 2000) seems more applicable.

If the users are not native English speakers provide the exact translations and definitions/descriptions of the terms used in the questionnaires. People from different cultures could understand the same terms differently. For example, words such as 'reliability', 'accuracy', and even 'trust' might mean different things to different nations (EATMP, 2003c).

3.4. Relation between trust and interface usability

The design rules and guidelines proposed above imply that there could be connections between the psychology of trust and the usability of HCI. Connections between certain usability qualities and trust have been demonstrated in a few recent works. For example the experimental study by Kim and Moon (1998) shows that some interface design features, such as cool color tones and balanced layout, can enhance user's perception of interface trustworthiness. Another example is the impact of the machine etiquette on user's trust (Parasuraman and Miller, 2004). However, this aspect of usability features is not yet studied sufficiently. In particular, the current studies provide only little experimental evidence, as yet, of the question concerning the different degrees of impact of different interface qualities on trust.

In this work we are also trying to address, at least partially, this important problem.

CHAPTER 4

IMPACT OF INTERFACE QUALITIES ON THE DEVELOPMENT OF TRUST

In this chapter we present experimental results on the impact of interface qualities on the development of trust. Here we investigate the hypothesis that the enhancement of the interface usability supports the development of trust.

To study this hypothesis we have conducted the usability evaluation of a novel remote sensing application IDFS (Intelligent Data Fusion System) developed by Lockheed Martin Canada (LM Canada). This chapter discusses the appropriateness of the application for this experiment. It provides a brief description of application, the selection of usability methods, and the steps to be followed for the evaluation of the IDFS interface usability. It presents the evaluation results and the recommendations for improving the interface usability. It also presents the results of trust measurements for both the old and the new (redesigned) versions of the interface for the task "Define mass functions". For measuring trust the questionnaire of Jian et al. (2000) has been applied. We conclude the chapter by discussing the relations between certain usability qualities and trust, and the appropriateness of Jian et al.'s (2000) scale for trust measurement.

4.1 IDFS application

The reasons for choosing IDFS application for our experimental study have been the following:

- IDFS is a novel application which was still in the process of development during this study. It implements recent technological developments in the domains of remote sensing and Data Fusion (DF) technologies that have not been implemented so far in other similar applications. This software offers functionalities for the computation of new features developed only recently, and it uses terminology that is commonly used in the DF community. Information is

provided through automated algorithms and processes. It has an associated level of confidence or uncertainty, which is not the case with similar applications so far. Because the application was new, and in particular it was implementing novel technologies, users did not have any prior experience in utilization of the application. So this software would be well suited for measuring the initial trust in the system.

- The priority task for R&D group of LM Canada, where the IDFS application has been developed, is the development and implementation of new technologies. The IDFS has also never been evaluated with respect of usability of its HCI and it has never been tested by real users. Thus, by evaluating the interface usability (where the target users will be involved as well) we could propose recommendations to enhance the interface qualities that would help to develop appropriate level of trust.
- Finally, since the software was still in the process of development, and since the developers at LM Canada could implement the proposed recommendations in the software, this system could give us the opportunity to check the effectiveness of the proposed recommendations and to test the hypothesis by measuring the operators' trust for both versions of the system, i.e. before and after usability enhancement.

4.2 General description of IDFS

The application used for our study, the IDFS, includes a number of new tools for image processing. This is an expert, or knowledge-based, system that implements recently developed technologies for image interpretation and DF processes. The main goal of the IDFS is to support an operator in target detection and interpretation. It provides a possibility to test several fusion schemes intended to give more reliable results for image interpretation. The built-in feature extractors and rule-based reasoning libraries make the IDFS suitable for various applications, such as military and intelligence operations, environmental and human activity monitoring, urban and coastal mapping.

According to software developers this application will provide an operator with a more comprehensive description of the observed scene, more reliable and quick information to support the operator in his/her decision making. It helps to identify the objects by fusing different information available about the object. The application is based on object-based image analyses, where additional information, such as shape, texture, area, and context can be used in the classification process.

The IDFS application applies feature-based fusion, i.e. first of all different features are computed, and then the fusion process is performed. The result is an image of fused features with soft decisions about the presence of different propositions. This feature-based DF incorporates the first level of the JDL model (See Appendix A), namely object recognition or refinement.

Another novel approach in this application is the fusion of the contextual information, i.e. information concerning relations between the objects. The user can define the mutual relationships between various objects in the present scene. For each detected object the contextual relation can either increase or decrease the confidence in the presence of that object. Spatial relations between objects can support the operator in cases of ambiguous interpretation, such as when two or more objects are related to the same region. This is already the second level of JDL (see Appendix A) model that helps to refine the situation.

Thus, the DF processes implemented in the IDFS consists of the following:

- Fusion of different features of radar sensors;
- Fusion of different features of optical (hyperspectral or ikonos images) sensors;
- Fusion of different features received from both radar and optical sensors;
- Fusion of different features and the user-defined context, i.e. the relationship between the objects in the environment both for radar and optical images.

Although the application can work as an automated system, i.e. the image interpretation can be carried out automatically, it also enables the user to carry out certain tasks manually, and in some cases the user can modify the already defined default automation.

4.3 Steps for the experimental study

In order to investigate the relations between the interface usability and trust, we need to enhance usability and to measure trust.

Two methods have been chosen for usability evaluation: heuristic and walkthrough evaluation. Heuristic evaluation is a good method for finding both major and minor problems in user interface based on established usability principles. However, no real user is involved in heuristic evaluation process. The advantage of a walkthrough evaluation is that actual users are involved in the evaluation process. It consists in observing the user's actions with the system, and interviewing users about their expectations and opinions about the software.

The next step will be to join in a single list the usability problems found from both the heuristic and the walkthrough evaluations by different users. This helps us avoiding repetition of the problems found (if the same problem is found by both methods) and to reveal the final list of the usability problems.

Before conducting the usability evaluation the following steps have been undertaken:

- Familiarization (of the author of the study) with the goals and operational functionalities of the software.
- Comparison of the new application with similar remote sensing applications widely used in the market, particularly in North America.
- Definition of target users.
- Definition of the goal driven task hierarchy, its validation by developers (see Appendix D).
- Definition of scenarios (validated with the R&D team) that include 80% of the

functionalities of the application (in the order of task hierarchy described above).

The usability evaluation of IDFS included the following:

- Heuristic evaluation of the HCI, with particular attention to the qualities that support the development of operators trust.
- Evaluation of the HCI with real users by the walkthrough evaluation method.
- Development of recommendations (based on two evaluation methods) for improving the usability and support for trust tuning via the HCI.
- Discussion of the feasibility of the recommendations with the R&D group.
- Changes in the HCI based on the proposed recommendations.
- Measurement of trust and performance for the old and the modified (improved) versions of the HCI.

The process of evaluation of trust included the following:

- Selection of users working or studying in the domain of remote sensing.
- Selection of trust measurement scale.
- Explaining to the users the objectives and the concepts of the software.
- Measurements of the task performance by the users.
- Rating of trust by the users on the basis of the Jian et al. scale (2000.).
- Calculation of values of separate trust items of the scale and of the overall trust.

4.3.1 Objectives of IDFS

In order to prepare the task scenarios it was necessary to define the objectives of the application and the sequence of tasks to achieve these objectives. Therefore all the available descriptions and documentation concerning the application have been studied (there is no user guide for the application yet). Besides regular meetings have been conducted with the R&D group, where the task hierarchy and the task scenarios have been validated, and different software functionalities have been demonstrated by the developers.

The IDFS application permits to achieve the following main objectives:

- Image interpretation of radar images;
- Image interpretation of hyperspectral images;
- Image interpretation of fusion of image interpretations of radar and ikonos images.

Appendix D shows the sequence of tasks to achieve the objectives (for all three interpretation cases).

4.3.2 Main tasks and their descriptions

For conducting the walkthrough and the heuristic evaluations, 23 task scenarios have been prepared (see Appendix D). The scenarios include the interpretation of all three types of images: radar, hyperspectral and ikonos. They not only include the main tasks to achieve the image interpretation but also the additional tasks that can help the operator to make final decisions.

In order to establish which problems should be fixed in the first place, for each task a priority scale (high, medium and low) has been established that took into account the importance of the task's objective (Table 4.1).

Table 4.1: The tasks and their priorities to achieve the objectives

#	Task	Priority	Image type
1	Create Region of Interest	High	All 3 image types
2	Extract features for the selected ROI	High	All 3 image types
3	Enter propositions for the selected ROI	High	All 3 image types
4	Define mass functions	High	All 3 image types
5	Fuse the mass functions of features	High	All 3 image types
5	Define contextual relations	Medium/low	All 3 image types
6	Get target information	Medium/low	Radar images
7	Get segmentation for several features	High	All 3 image types
8	Get segmentation for one feature	Low	All 3 image types
9	Get the pixel distribution in 3D plot	Low	All 3 image types
10	Get polarimetric synthesis	Low	Radar
11	Browse through hyperspectral window and get the spectral signature	Low	Hyperspectral
12	Identification of endmembers' names	Medium	Hyperspectral
13	Perform target detection	High	Radar, Ikonos
14	Get the image interpretation	High	All 3 image types
15	Fuse the multisensor interpretation	High	Radar, Ikonos

- Tasks classified as "high" priority are the tasks that MUST be carried out in order to reach the goal.
- Tasks classified as "medium" priority are the tasks that should be done to get more refined result however without this task the user can reach the goal.
- Tasks classified as "low" priority are the tasks that can be carried out by the operator to get additional information for image interpretation.

4.3.3 Target users and comparative analyses of the software

In order to conduct the walkthrough evaluation we identified the potential target users of the application IDFS. The following criteria were taken into consideration:

- Users should have experience with currently most popular remote sensing applications in North America;
- It is important to find an expert user in radar and optical image interpretation.

The research revealed that the most popular and used remote sensing applications in North America are the applications PCI and ENVI. Hence, the walkthrough evaluation has been carried out with users that currently work with PCI, have worked with ENVI, and are experts in radar or optical or in both image interpretation.

Table 4.2 shows the users' experience with the software they use in their work.

Table 4.2: The users' experience with the domain

Users	Experience
User 1	8 years experience working with PCI and ENVI (mostly working with optical images)
User 2	7 years experience working with ENVI , PCI, Grass (working with all kind of images, but presently with radar images)
User 3	12 years experience with PCI and ENVI (working with all kind of images)
User 4	6 years experience working with PCI (specialized in radar images)
User 5	14 years experience (working with all kind of images)
User 6	15 years experience in remote sensing

Comparing the IDFS with similar existing applications helps to define the similarities and differences of the concepts and approaches of the new application and the ones already in use. Since the IDFS is developed in Canada and is aimed at the North American market, in the selection process particular emphasis has been given to applications used in that market.

The consultations with the users (from companies "Tecsult", "ViaSat geo Technologies", "Laser Map"), professors and graduate students (from the Department of Geography of the University of Montreal) who work in the area of remote sensing, have revealed that the most utilized remote sensing software in North American market are PCI (in Canada) and ENVI (in USA). Identification of the objectives of these applications, familiarization with them, and meeting with the users revealed that the basic concepts implemented in the IDFS application are significantly different from the ones used in PCI or ENVI.

Below are listed the new concepts used in IDFS that are not found in PCI or ENVI:

- Region or object based segmentation.
- Feature and information based fusion.
- Fusion of contextual relations between the objects.
- Presentation of results by soft decisions.
- Concept of mass function (the weight of the belief of presence of certain proposition).
- Automated tools, with possibilities of their modifications, to conduct the image interpretation.
- Weighting (by the user) of different computed features in the process of fusion.
- Computation of certain recently developed features.
- Utilisation of terminology common in the DF community.

Some of the features of the software (e.g. definition of mass functions, presentation of results by soft decisions) are also found in the German software E-Cognition. However, the E-cognition is unknown in Canadian market. There was only one user, quite interested in recent developments for his research and who had some (rather vague) information about E-cognition application.

4.4 Heuristic and walkthrough evaluation

Before conducting the walkthrough evaluation with users we have conducted a heuristic evaluation of the HCI for all tasks presented in Table 4-3 using 23 tasks scenarios described in Appendix C. The evaluation has been carried out (by the author of the current memoir ho has already applied this evaluation method in her Master's projects) based on the usability criteria (see Appendix E) proposed by Robert (1997).

The walkthrough evaluation has been conducted with six expert users presently working in the domain of remote sensing (see Section 4.3.3). All the users have had from 6 to 10 years of working experience in this field. Before starting the evaluation the goals and the tasks of the new application have been explained to the users. Then the users have performed the tasks suggested in the scenarios (see Appendix C). The tasks of high priority described in Section 4.3.2 have been performed first, using mainly the radar imagery. Then, depending on the time left, other tasks, including the ones with hyperspectral and Ikonos-radar images, have also been performed. In the process of conducting the tasks the users were commenting their work. They were also answering the questions concerning the new tools, the new functionalities, and the problems they have been facing. Depending on the situation some of the tests have been recorded on the videotape or the audiotape, and in some cases we were simply taking notes.

The problems that were found have been summarized in separate lists for each evaluation method. Then we merged the two lists together eliminating the repetition of problems revealed by both evaluation methods. We also proposed possible solutions for each problem. The solutions were based on the checklist of Robert (1997), on guidelines for designing user interface (Smith and Mosier, 1986), on trust related guidelines proposed in this study, and on the expert users' opinions collected during the walkthrough evaluation.

The final list of the usability problems found with the heuristic and walkthrough evaluations is presented in Table 4.3. It shows the usability problems for each task, the priority of the task, and also the expected gravity of the problem for reaching the goal.

The two evaluation methods revealed a total of 46 problems. These included 20 high gravity problems (18 problems found by heuristic evaluation and 14 found by walkthrough, many of which were revealed by both methods), 19 medium priority problems (13 found by heuristic evaluation and 13 found by walkthrough), and seven low priority problems (5 found by heuristic evaluation and 3 found by walkthrough).

Based on the guidelines proposed in this study (see Section 3.3) we determined whether the usability problems could have an impact on the user's trust or not. The last column of the Table 4-3 indicates possible impact on trust for each of the problem.
The acronyms presented in the table are the following:
EM: Evaluation method;
PG: Problem gravity: h = high; m = medium; l = low;
TP: Task priority: h = high; m = medium; l = low;
TI: Trust impact;
PF: Problem is fixed;
H: Heuristic method
W: Walkthrough method

Table 4.3: The list of problems revealed by heuristic and walkthrough evaluation

Task	Usability problems	Possible solution	EM	PG	TP	TI	PF
Open file/ Create ROI	The form of the selected ROI (by user) and displayed ROI are not compatible.	Keep the same form (unit scale) for the both ROI.	H W	h	h	Yes	Yes
	Three clicks are required to conduct the task.	Provide an icon for ROI on tool bar.	H W	h	h	No	Yes
	Only rectangular form is available.	Provide other forms, if feasible.	H W	l	h	No	No
	While cancelling the opened window "Open HDF5 file", the window disappears, but the file opened previously disappears as well.	Fix the software problems.	H	m	h	Yes	No
Create an independent region	While selecting the "Create Independent Region", the displayed image (e.g. ikonos) disappears and the other image (e.g. radar) appears. This is inconsistent with the task "Create Region" and with the user expectations.	Provide the user with a prompt how to create 2 images in one file or to provide on separate windows.	H W	l	h	Yes	Yes
Select ROI	For HSI-fusion task there is no prompt "Select ROI".	Fix the problem in software	H	l	h	No	No
	When selecting a new ROI, the old ROI image remains on display.	When new ROI is highlighted to activate that window.	H	h	h	No	Yes

Table 4.3: The list of problems revealed by heuristic and walkthrough evaluation (continued)

Select ROI (cont'd)	Three clicks are needed to select ROI.	Provide icon on "IDFS" window.	H W	l	h	No	Yes
Extract features	Some new feature names are not familiar to user.	Give a short description.	W	m	h	Yes	No
	There is no feedback for already computed features; the feature is being computed again.	Give a prompt.	H W	m	h	Yes	No
	Three clicks are required to compute a feature.	Provide icon on "IDFS" window.	H W	m	h	No	No
	It is not clear what data is entered for CFAR feature extraction.	Provide the history	H W	m	m	Yes	No
	There is an order to compute certain features, else the system aborts: -For radar images: the features "Backscatter" and "Amplitude" should be computed in the first place. -For Ikonos: in the first place "Band Sharpening" feature should be computed.	Provide these features in a right order in the menu of features. In the case of a wrong order selection provide a prompt pointing the features to be computed	H W	h	h	Yes	No
	If ROI is too big, system cannot calculate the feature "Edge detection". However, after the command the system starts the calculation and then aborts. This damage the image file.	Stop the calculations if the limits are exceeded. At the same time to give the user an explanation by a prompt	H	h	m	Yes	No

Table 4.3: The list of problems revealed by heuristic and walkthrough evaluation (continued)

Extract features (cont'd)	In certain cases there is no consistency between the feature name and the name appeared on HDF browser, particularly the features: For radar images		Keep consistency, e.g. keep the same name.	H	h	h	Yes	Yes
	Backscat-tering coefficient	Back-scatter						
	PWF	Filtered SAR						
	Cameron CTD	Cameron						
	Compute Polarization Signature	Polarimet-ric Response						
	Cloude's Decomposit ion and Cloude Classifier	Cloude						
	Subaparture Coherence	Sub-aparture Coherence Azimuth Right Azimuth Left						
	Edge Extraction	Edges Detection						
	Ratio Line Extraction CC Line Extraction	Line Detection						
	For ikonos images							
	Band Sharpening	PS_MS						

Table 4.3: The list of problems revealed by heuristic and walkthrough evaluation (continued)

Target informa-tion	For certain images it provides a black window without any information (e.g. if the image quality is not good).	To give the reason by a prompt.	H	m	m	Yes	No
Target information (cont'd)	The meaning of the provided parameters is not clear	Provide a "quick help" icon for explanation	H W	m	m	Yes	No
Define mass functions	Terminology is not familiar.	Provide more familiar terminology or an explanation	H W	h	h	Yes	No
	The graphical display is not clear: no values, no units, and no borders are present.	Provide corresponding units for the pixel values and the confidence. Provide borders.	H W	h	h	Yes	Yes
	The belonging of mass functions to propositions is not clear. There is no traceability for the belonging of the mass function graphs.	Provide the name of the proposition on the graph of the mass function.	H W	h	h	Yes	Yes
	System does not always provide a feedback while the user draws the mass function	Provide an explicit feedback for each entered point (i.e. click).	H W	h	h	Yes	Yes
	The confidence and pixel values are in pale color. Hard to read the values.	Provide the values in black.	H W	h	h	No	Yes

Table 4.3: The list of problems revealed by heuristic and walkthrough evaluation (continued)

Define mass functions (cont'd)	Graphs for all propositions are in the same colour. The frequencies (when using that method) have the same color as the drawn mass functions.	Provide each graph of mass function in different windows. For the task "Show Frequencies" provide the frequency graph in different color than the mass functions graph, or show it on another window.	H W	h	h	Yes	Yes
Enter proposition	Terminology is not familiar.	Provide an explanation	W	h	h	No	No
Segmenta-tion of different features	The results are not understandable, particularly the shape parameters.	Provide 'help' on the window for quick access for an explanation.	W	m	h	No	No
	Terminology "Growing region" in the menu is not familiar to the user. It has different meaning for PCI and ENVI. Besides, the segmentation result appeared on the HDF file appears as "Segmentation_SAR / Ikonos / Hyper (according to type of image)" and instead of "Growing Region".	Use familiar and consistent terminology, i.e. the Segmentation_ SAR and provide an explanation of the terminology.	W	m	h	No	No

Table 4.3: The list of problems revealed by heuristic and walkthrough evaluation (continued)

Segmenta-tion of different features (cont'd)	When the window "Region Growing Config." is opened the command "Cancel" causes a system abortion.	Fix the software problem	H W	h	h	Yes	No
Interpret image	There is no "Undo" possibility once the window of "Segmen-tation level or feature selection" is opened. Any click on the "Cancel" causes abortion of system.	Fix the software problem.	H W	h	h	Yes	No
Interpret image (cont'd)	While selecting the segmentation level: both the SAR and Ikonos levels (if both are computed) are displayed on the window without any indication which level belongs to what type of image. The same is for selecting the features.	Display the "Level" for each image under corresponding section (e.g. section "SAR"; section "Ikonos").	H W	h	h	Yes	Yes
	It's not clear what features are entered for the fusion result.	Provide the history.	H W	h	h	Yes	No
Target detection	If the window for "Fuse detection" is opened, in case of "Cancel", the system aborts.	Fix the software problem.	H	h	m	Yes	No
Endmember identifica-tion	"Fraction Map" terminology is not clear.	Provide its explanation.	W	m	m	No	No

Table 4.3: The list of problems revealed by heuristic and walkthrough evaluation (continued)

3D plot	The objectives of the commands "Analyse segmentation's cluster", "Analyse hyperspectral cluster" and "Show Cloud's scattering zone" are not clear.	Provide 'help' on the window for quick access for a description of each action.	W	m	m	No	No
	For "Show cluster" the number of cluster can be entered only via keyboard.	Provide a menu for number selection.	H	l	m	No	No
	There are no dimensions on axes.	Add a scale (dimensions)	H W	m	m	No	No
Segmentation of one feature	The terminology "Fingerprints" is not familiar.	Provide an explanation.	W	m	m	No	No
Segmentation of one feature (cont'd)	There is no indication about the number of segments used.	Provide the history.	H	l	m	No	No
Hyper spectral window	After selection of other ROI, the old ROI does not disappear. Hence, the user needs to close the old ROI and open the new selected ROI.	Enable the system to give the new ROI without all these steps.	H	m	l	No	No
Polarimetric window	For Y axes there are no normalised values.	Give normalised values on Y axes (as in software PWS).	W	l	l	No	No

Table 4.3: The list of problems revealed by heuristic and walkthrough evaluation (continued)

General	In the fused radar/ikonos image it is not clear what soft decision belongs to what proposition.	Provide a possibility to get the beliefs for that pixel while clicking on that pixel.	H W	m	h	No	No
	No zoom is possible.	Provide zoom	H W	m	h	No	Yes
	Only one window at a time is available.	Provide as much windows as the user needs.	H W	m	h	No	Yes
	All the windows have no indication.	Provide the corresponding indication on top of the window.	H W	h	h	Yes	Yes

4.5 Interface redesign

The usability problems that were found and their possible solutions (see Table 4-3) have been discussed with the developers of IDFS in the R&D group of LM Canada. Since it was problematic to implement all these solutions at once, the attention has been focused on the interfaces of the tasks that are the most important for achieving the automation goals and that could have an impact on trust. In this respect the highest priority was given to the task "Define mass functions". The reasons for this are the following:

- The entire fusion process and the final image interpretation depend on the results of this task. The correctness of determination and entry of this parameter by the operator is very important for the final result.
- The results of the interface evaluation show that "Define mass functions" is perceived as one of the most difficult tasks for the users. This is a new task wherein the operator manipulates a concept that is unfamiliar for all users participated in the IDFS evaluation. The concept of mass function is utilized neither in PCI nor in ENVI. Besides, in the initial version of the IDFS the interface design for this task was very poor.
- The evaluation has revealed a number of usability problems related to this task (listed in Table 4-3) that could have an impact on the operators' trust in the system.

The initial version of the interface for the task "Define mass functions" is shown in Figure 1. Because of its importance, we describe below in more details this task and the improvements implemented in the new interface.

In this task the user defines the mass functions for different propositions which help to perform segmentation of the image (for more details see Appendix C, Scenario 6). For a given proposition the user needs to check the reflection values[4] on the image of the computed feature and to enter the mass function on the window "Define mass functions"

[4] The reflectivity coefficient of the objects in the scene.

by drawing the graphs about the presence of selected proposition for the range of values of interest for that proposition. The horizontal axis on the window shown in Figure 1 corresponds to the range of reflectance values, and the vertical axis shows the confidence level of the operator concerning the presence of that proposition at the given reflection value. This entered mass function is used in the next steps for the final image interpretation (see Appendix E). Thus, this entry is very important for the end result of data fusion and image interpretation.

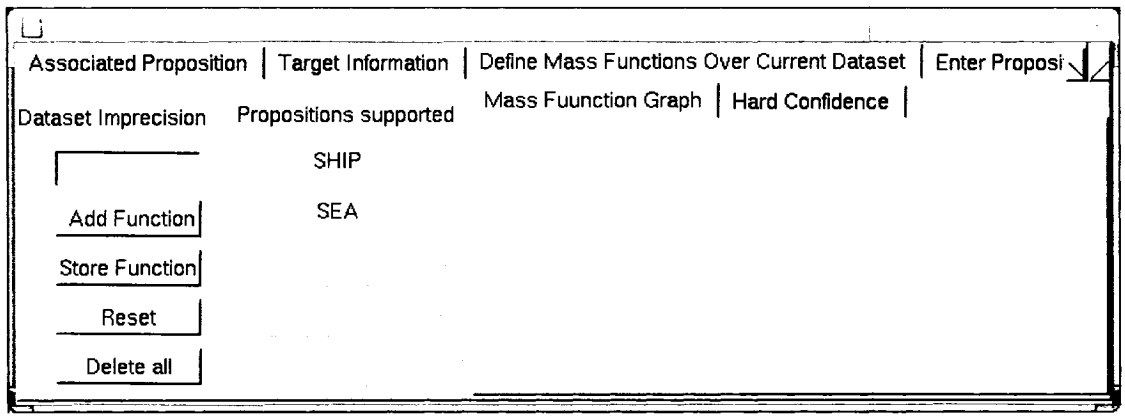


Figure 4.1: The old version of the "Define mass functions" interface.

The usability problems found in the initial interface of the task "Define mass functions" are the following:

- As we can see in Figure 4.1, the interface does not provide any distinct frames, labels of the axes, digits that show the range of variation of the feature parameter, so it was problematic for the user to perform correctly the task (visual clarity).
- The color chosen for drawing/writing in the window was pale (yellow) as shown in Figure 4.1, and for most users it was difficult even to read the reflection values in the window(visual clarity).
- If the region of interest (ROI) contains several propositions (i.e. several classes) the user needs to draw the mass functions for all these propositions in the same window using the same color (in Figure 4.1 we see two mass functions for two propositions). Therefore such interface did not provide an easy traceability, i.e. it

was difficult to visually distinguish between different mass functions for different propositions. When analysing several propositions it was becoming particularly messy (traceability of data).

- The system did not provide a good feedback for the data entry. There was no feedback on mouse clicks on the window to show the pointer coordinates (Data entry; informative feedback, compatibility with user's expectations).
- If the user needed to make changes of the entered values, it was not evident how to delete the previous value and enter the new one (data entry, visual clarity).
- The users prefer to have some preset versions for data entry (which is the case with the software MATLAB). This gives more flexibility for drawing the graphs and more guidance for the users to draw functions (compatibility with user's expectations).

The developers of IDFS have modified the interface for the task "Define mass functions", making it more similar to the interface of the known software MATLAB where a concept similar to mass function is implemented. Five problems out of the six found have been fixed. According to the guidelines discussed above, four out of these five problems could have an impact on trust (see Table 4-3). The improved version of the interface of the task "Define mass functions" is demonstrated in Figure 4-2.

The following problems have been fixed:

- The new interface provides frames, labels on the axes, and the range of reflectance values being analyzed (visual clarity).
- The reflection values are now displayed in black color, and the mouse coordinates are displayed both at the pointer position and separately below the frame (visual clarity).
- The interface provides an easy traceability for each mass function entered. Now each mass function is drawn on a separate frame and one can navigate through these frames by selecting different numbers on "Display function". Each frame displays its proposition name (in the menu "Proposition supported") and the defined mass function (traceability of data).

- The system provides a good feedback for the entered data; the user can always see the entered node and the lines between the nodes (data entry; informative feedback, compatibility with user's expectations).
- The user can now make changes for each entered value, for each node of the preset functions by pressing "Delete the node". The user can also add a node at any time by pressing button "Add node" (data entry, visual clarity).
- The system provides more flexibility for drawing mass functions. It now provides preset versions for mass functions, as it is the case with MATLAB software. Therefore the new interface will be more familiar for users that already have an experience with the MATLAB (compatibility with user's expectations).

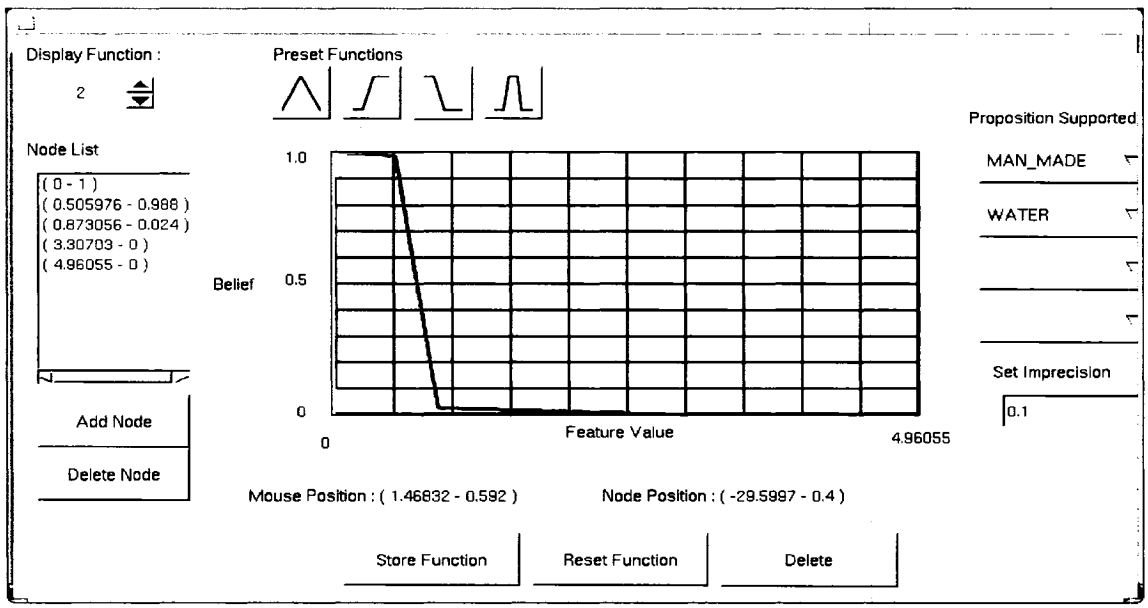


Figure 4.2: The new version of the "Define mass functions" interface.

The second column of Table 4.3 shows the usability problems, and indicates which problems have been fixed. Although the developers have also fixed a number of usability problems besides the "Define mass function" interface, here we do not discuss these improvements since the focus of this work is on the usability features that would have an impact on users' trust.

4.6 Trust measurement

In this section we present the methodology and the results of trust measurement for the initial and the new (redesigned) versions of the "Define mass functions" interface. It was expected that the overall trust in the IDFS with the improved interface would be enhanced compared with the initial version of the system. We also expect that the new interface will enhance the users' performance. Therefore we measured the performance time for the "Define mass function" task in order to see whether this expectation between the performance and trust is correct.

4.6.1 Methodology

Participants. Six graduate students (1 woman and 5 men) doing research in the domain of remote sensing in the Department of Geography at the University of Montreal have volunteered to participate in the evaluation test. The age of participants varied from 25 to 35 years. All six users were French speakers; however, all of them could speak and read English as well.

Material. The tests have been conducted on a Pentium IV laptop with a 17" high resolution (1024 x 768) monitor.

Procedure. The participants performed the task "Define mass functions" with the old and the new interface versions of IDFS. We standardized the procedure of the tests as much as possible. All tests have been conducted separately and at the same location (remote sensing laboratory in the Department of Geography). The duration of each test was approximately one hour. Before conducting the test we explained to the participants the purpose of the software and described the new features used in the software.

In order to avoid user biases we tried to be as neutral as possible; we did not give any preference to the new or the old interface versions and asked the users to be as objective as possible while evaluating. We mentioned to them that the software was still in development and that we were open to their comments or opinion concerning any future

changes for both interface versions. First we demonstrated how to conduct the task for the new and the old interface versions (see scenario 6 in Appendix C.1 and Appendix C.2), then the task was performed by the participants in order to allow some familiarization with the software. After the participants acquired some experience with conducting the task, the performance time was measured with a stopwatch. The author of this study observed the users during the tests. The testing order of the interface was the same for all users: they tested the new version first and then the old one. A possible solution for avoiding any bias could have been a mixed order of testing for the two versions by different users. However, we decided always to start with the new interface version first. Walkthrough evaluation conducted earlier has revealed that the users had difficulties with the old interface version. The difficulties were due to poor usability of the interface. Since the new version of the interface provided a significantly better usability, it would be easier to conduct the task with this version first, which would then facilitate conducting the task for the old interface version.

We assume that this order provides some advantage to the old interface version for the following reasons. Users perform the same task with both interface versions wherein the concept of mass function is used. The difference in usability features lays in a better visual clarity, feedback, traceability and flexibility of the new interface version. The new interface version provides more transparency on how to conduct the task and gives more possibilities for data entry, as described in Section 4.5. Hence we assumed that testing of the new version of the interface allowed the users to acquire some experience with the task and the system and helped them to work with the old interface. After conducting the tasks the participants completed the measurement scale for each version.

Selection of trust measurement scale. The selection criteria for trust measurement scale were the following:

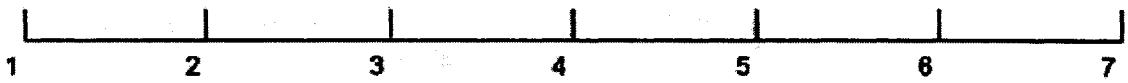
- Scale that is based on experimental studies.

- A multiple rating scale, including several questions. This will help to reduce errors in the final result due to uncertainties or insufficient confidence in users' answer (note that this is a subjective rating).
- A generic that is not restricted to a specific domain.
- A scale that allows us to measure the initial trust, when a user has no prior experience with the system.

We can see from the table 3.1 (see Section 3.1.1) that the measurement scale of Jian et al. (2000) satisfies all criteria mentioned above. The scale of Jian et al. (2000) is shown in Figure 4.3.

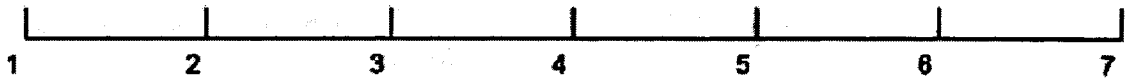
Please mark an 'x' on each line at the point which best describes your feeling or your impression. (Note: 'not at all' = 1; 'extremely = 7)

1 The system is deceptive.



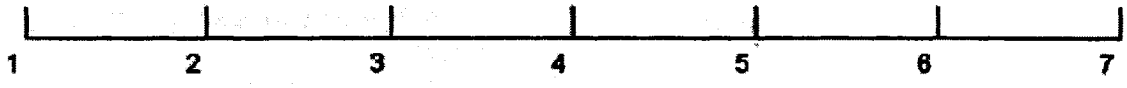
2. The system behaves in underhanded manner.

3. I am suspicious of the system's intend, action, or outputs.



4. I am wary of the system.

5. The system's actions will have a harmful or injurious outcome.

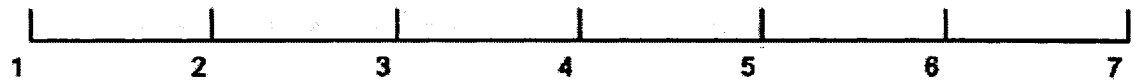


6. I am confident in the system.

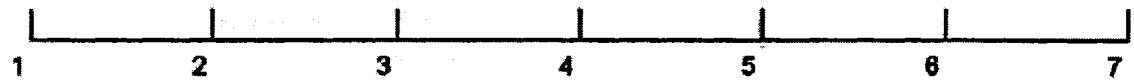
7. The system provides security.

8. The system has integrity.

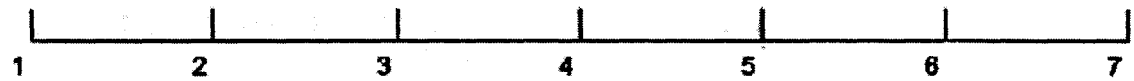
9. The system is dependable.



10. The system is reliable



11. I can trust the system.



12. I am familiar with the system.

Figure 4.3: Trust measurement scale of Jian et al. (2000)

We have translated in French the measurement scale of Jian et al. (2000). It was checked by two native French speaking scientists with a Ph.D. degree working in the R&D group at LM Canada.

The French translation of Jian et al. (2000) the measurement scale

1. The system is deceptive.
   Le système est trompeur.
2. The system behaves in underhanded manner.
   Le système se comporte d'une manière trompeuse.
3. I am suspicious of the system's intend, action, or outputs.
   J'ai des doutes sur l'intention, l'action ou les outputs (sorties) du système.
4. I am wary of the system.
   Je suis prudent avec le système.
5. The system's actions will have a harmful or injurious outcome.
   Les actions du système peuvent avoir des conséquences/résultats nuisibles.
6. I am confident in the system.
   Je me fie au system.
7. The system provides security.
   Le système fournit de la sécurité.
8. The system has integrity.
   Le système a de l'intégrité.
9. The system is dependable.
   Je peut compter sur le système.
10. The system is reliable.
    Le système est fiable.
11. I can trust the system.
    Je peux faire confiance au système.
12. I am familiar with the system.
    Je suis familier avec le système.

The participants have been told that they could skip any question if they did not understand it or could not answer it. The measurement scale does not provide definition of the terms or an explanation of each item. We have been anticipating that the users could ask questions on certain terms; therefore, we prepared the definitions for potentially ambiguous terms in the questionnaire, like "Integrity", "Reliability" and "Dependability". These definitions have been given to the users along with the questionnaire. The definitions of the terms used in the trust measurement scale have been equally explained to each user.

Definitions for certain terms of the scale (presented the users)

Integrity: The system adheres to a set of principles that the operator finds acceptable (extracted from Lee and See, 2004).
Reliability: The extend to which you can count on the machine to provide appropriate support to the tasks (from Taylor et al., 1995).
Dependability: The extend to which you can rely on the machine to consistently support the tasks (from Taylor et al., 1995).
Confidence: Confidence in ability of the machine to support successful completion of the tasks (from Taylor et al., 1995).

The scale of Jian et al. (2000) allows one to measure various trust-building parameters. However, it does not provide measuring of the user's overall trust which would be useful. So we measured the overall trust by applying the approach proposed in Section 3.2 for estimation of the overall trust.

4.6.2. Results

The results of the questionnaire are presented in Table 4.4 and Table 4.5. The last item of the questionnaire ("I am familiar with the system") has been discarded as irrelevant and misleading since all users understood it as literally questioning their previous experience with the IDFS. The numbers on an item in both tables is consistent with the number of an item provided in Figure 4.3.

The questionnaire uses a seven point rating scale with possible answers labelled from "not at all" to "extremely". The ratings given by six users for each question are presented in the corresponding columns. For each item, I_i represents the mean value of the ratings of six users for the i^{th} item.

Table 4.4: Results of trust measurements for the old interface

Item #	User 1	User 2	User 3	User 4	User 5	User 6	Mean $\mathbf{I_i}$	Norm $\mathbf{P_i}$	Adjusted value $\mathbf{W_i}$
1	6	3	6	4	5	5	4.833	0.639	**0.361**
2	6	2	5	7	5	2	4.500	0.583	**0.417**
3	7	3	5	5	2	5	4.500	0.583	**0.417**
4	7	3	2	7	7	7	5.500	0.750	**0.250**
5	7	N/A	N/A	4	1	4	4.000	0.500	**0.500**
6	1	5	3	1	3	2	2.500	0.250	**0.250**
7	1	6	3	2	4	2	3.000	0.333	**0.333**
8	1	4	6	3	6	2	3.667	0.444	**0.444**
9	1	5	5	1	6	3	3.500	0.417	**0.417**
10	1	5	5	1	2	5	3.167	0.361	**0.361**
11	1	5	4	1	5	4	3.333	0.389	**0.389**
Overall (averaged) trust T									**0.376**
Standard deviation σ									**0.077**

As discussed in the Section 3.2, it is usually more convenient to express the results of measurements in percents rather than in the units that can vary from 1 to 7. The procedure for such normalization has been explained in Section 3.2 above. In Tables 4.4 and 4.5, P_i is the normalized mean value for the *i*-th item. It is obtained by a simple formula $P_i = (I_i - 1)/6$. In normalized units the answer "not at all" corresponds to 0, and the answer "extremely" corresponds to the value 1 (i.e. 100% trust). W_i is the adjusted trust value which takes into account whether the value of the i^{th} item has negative or positive meaning for trust. Thus, $W_i = P_i$ for positively trust-related items and $W_i = (1-P_i)$ for negatively trust-related items. T is the overall trust value. The N/A corresponds to questions skipped by some users (the cases when the users could not understand or

answer the question). And the last column shows the average values for each item of the measurement scale. The last two lines show the overall trust and the dispersion value ("standard deviation") of all 11 items from overall trust.

Table 4.5: Results of trust measurements for the new interface

Item #	User 1	User 2	User 3	User 4	User 5	User 6	Mean $\mathbf{I_i}$	Norm $\mathbf{P_i}$	Adjusted value $\mathbf{W_i}$
1	1	2	3	4	1	2	2.167	0.194	**0.806**
2	1	2	4	2	2	1	2.000	0.167	**0.833**
3	1	3	5	2	1	2	2.333	0.222	**0.778**
4	1	3	2	2	4	2	2.333	0.222	**0.778**
5	1	N/A	N/A	4	1	2	2.000	0.167	**0.833**
6	5	6	4	5	5	5	5.000	0.667	**0.667**
7	6	6	4	5	4	5	5.000	0.667	**0.667**
8	2	6	7	6	6	1	4.667	0.611	**0.611**
9	5	6	6	4	6	5	5.333	0.722	**0.722**
10	6	5	6	5	2	6	5.000	0.667	**0.667**
11	6	5	5	4	5	6	5.167	0.694	**0.694**
Overall (average) trust T									**0.732**
Standard deviation σ									**0.076**

The results of trust measurement (presented in Tables 4.4 and 4.5) are consistent with expectations for improved trust. Trust in the new interface has been increased by 2 times; it has enhanced from $T_{old} = 0.37$ to $T_{new} = 0.73$. The calculation of the dispersion σ (i.e. the square root of the variance) shows the effective width of the distribution of 11 individual trust values for the items in the measurement scale. The deviation of these values from the mean value T is small, $\sigma \approx 0.077$ for old and $\sigma \approx 0.076$ for the new versions of the interface; so we have $T_{old} = 0.376 \pm 0.077$ and $T_{new} = 0.732 \pm 0.076$. It shows that in both cases the deviation of individual trust parameters in the scale from the overall trust value is very small, i.e. the individual parameters in the scale almost equally reflect trust in the system. Tables 4.4 and 4.5 show that trust has been enhanced for all items of the scale.

Figure 4.3 shows the results of trust measurements (in percentage) for the old and new redesigned interface versions.

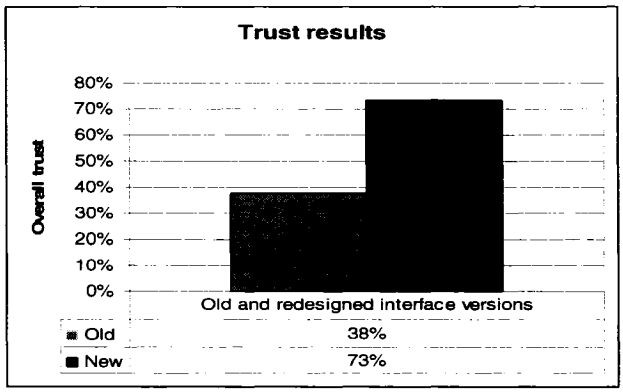


Figure 4.3: Overall trust (in %) for the old and the new interfaces.

While the users were performing the task "Draw mass functions" we also measured the performance time for that task for the old and new interface versions. Although this is not the objective of this study, we found it interesting to see the connections between the changes in performance time and changes in trust. The performance time has been measured with a stopwatch with all six subjects. As we can see from Table 4.6, the mean performance time for six subjects has been improved (i.e. decreased) by 36% (Reduction of performance time: $(T_{old}-T_{new})/T_{old}$).

Table 4-6: The performance time (in seconds) for the old and new interfaces

	User 1	User 2	User 3	User 4	User 5	User 6	Mean
Performance time (sec) for old interface	345	355	295	305	80	195	
Performance time (sec)for new interface	145	170	270	230	80	82	
Performance (speed) enhancement	2.38	2.09	1.09	1.33	1.00	2.38	1.71
Reduction of performance time	58%	52%	8%	25%	0%	58%	36%

The value for the factor of paired performance speed increase (i.e. if we calculate the performance speed for each user and then calculate the average for all these speeds) is

equal to 1.7. Paired *t*–test (a two-tailed *t*-test was used for paired samples) was conducted to determine if there is a significant difference in the users' performance time between old and new interface. It revealed a significant difference in performance time ($t = 2.976$; $p = 0.031$). However, it would be a bit premature to draw final conclusions based merely on this test. For more conclusive results a number of tests including a larger number of subjects, tests, and system functionalities is required.

4.6.3 Discussion

The experimental results show that the enhancement of the usability qualities of the user interface has increased trust in the system. The new interface offers higher precision for data entry, traceability of the data, a better visual clarity, feedback and compatibility with the users' expectations. The users did recognize that certain changes had no direct links with the system capabilities. Nevertheless, the results of the measurement scale have showed that the user's trust in the system increased from 38% to 73%. This corresponds to an enhancement of trust by a factor of 2. The deviation of the 11 trust-related items from the overall trust value is small, $\sigma \approx 0.077$ for the old version and $\sigma \approx 0.076$ for new interface. These small dispersion values imply small widths of corresponding distributions of trust values of individual items (Wi) around the overall trust. For the old and the new versions of the interface these values are effectively distributed in the regions of 0.38 ($\pm$ 0.077) and 0.73 ($\pm$ 0.076), respectively. The absence of any overlap between these two regions strongly supports the conclusion that trust has been notably improved.

The results are also consistent with the consideration of Lee and See (2004) and Miller (2005) that in interactions with a novel system both the affective and analogical trusts play an essential role. The enhancement of trust in the case of IDFS could not be attributed to the analytical trust developed by the users, since the users had no working experience with the system. It should be rather ascribed to affective (i.e. the first effective factor) and analogical trust (i.e. weather the system's behavior is consistent with user expectations relative to the rules and procedures). For example, the

walkthrough evaluation with expert users revealed that for the task "Draw mass functions" the users had expectations to see an interface that could support them to understand how to enter the data or to change data already entered, how to show labels on the axes, and an interface that would give a good feedback when the data is entered. Since these problems have been mostly fixed for the new interface, we can conclude that the enhancement of trust has been related to affective and analogical trust development as well. During the tests, while testing the old interface, the users were becoming frustrated and confused when the system did not give an informative feedback about the data entry.

Hence, we can conclude that the enhancement of usability qualities supports the affective and analogical trust development.

The measurement scale of Jian et al. (2000) allows trust rating by users without experience with the system. There seems to exist some redundancy in the items in the sense that the questions are correlated to some degree. Also, the scale embraces several items of trust, both negatively and positively connected with trust. Both these features, however, appear useful for a more correct measurement of the overall trust, because they help to reduce, by averaging, the impact of the user slips in ratings due to either unfamiliarity with the system, or with uncertainty in his/her answers.

The tests have also revealed the following problems related with the scale itself.

The scale does not provide definitions for the terms used. We consider this as a significant insufficiency because of the following reasons: Although the French translation of the questionnaire has been provided, all the users had noticeable difficulties to correctly understand the meaning of many terms. Examples are given by 'integrity', 'dependability', 'security', or by difficulties in understanding clearly the difference between the terms 'reliability' and 'trust' in the system. In some cases the meaning of the individual terms taken separately was clear to the users, but there were

difficulties with explicit understanding of some questions. For example, it was not clear what "The system provides security" exactly means.

Although users seem to understand the meaning of the questions, it could vary from one user to another. Even in research literature the same term may have different definitions. For example, the term "integrity" by Lee and See (2004) is defined as follows: "***Integrity*** is the degree to which the trustee adheres to a set of principles the truster finds acceptable", whereas according to Madsen and Gregor (2000) it is define as follows: "***Integrity*** means that the system is able to recover from technical failures or user errors without loss of data". Thus, the users can be misled in their perception of the right meaning of the questions.

It should be noted in this regard that the scale has been evaluated by seven professional linguists involved in the process of assessing the terms. As a result the questionnaire included 12 different trust-related items that imply good knowledge of English. Yet the operators are mostly people with very different linguistic knowledge and background.

As discussed in Section 2.3.2, trust is a multivariate construct. Therefore the presence of several items in the measurement scale is a requirement. However, the scale of Jian et al. (2000) does not suggest any means to unify the 12 answers in a single overall trust measure. It would be useful to have such a single overall trust result, as proposed in the Section 3.2 above. We also note that the proposed unification model does not eliminate the results on individual items. One can always use these results to estimate specific aspects of trust.

4.6.4 Limitations

The results presented in this study have some limitations.

Firstly, only one task, although the most important one, and only one set of functionalities in the system, have been tested by the users. Secondly, it would be preferable to have a larger number of subjects to measure trust. Unfortunately we had access to a limited number of subjects having experience with similar applications.

Another limitation is the lack of adequate descriptions of the terms used in the scale. In order to be consistent, for all users we provided the same definitions of terms found in relevant research literature. In some cases, when the appropriate interpretation could not be found in the literature, we proposed definitions after consulting with expert users. This was the case with the item "The system's actions will have a harmful or injurious outcome", for which we explained that it relates to the safety of the entered data, their possible damage or crash. In another case, the meaning of the term 'security' was incomprehensible for the users: it was not clear what kind of security it implied. Since we could not find any explicit explanation in the literature in this domain, we preferred to leave the users with the freedom of their own understanding of security.

In this study the weight of each item in the measurement scale is assumed to be equal. For both the old and new interfaces the dispersions ($\sigma_{old} \approx 0.077$ and $\sigma_{new} \approx 0.076$) of the responses for each trust item from the overall trust value is small. However, differences between the weights of these items cannot be excluded.

Finally, the order of testing the interface version was always the same (first the new interface and then the old interface version), which could create a bias. The justification for this order of testing is given in the Section 4.6.1.

CHAPTER 5

CONCLUSION

This work presented the state of the art on the concept of trust in complex automated systems. It proposed generic rules and guidelines for the design of automated systems on how to support an appropriate level of trust in systems, and particularly in novel ones. The rules and guidelines are based on theoretical, experimental and empirical results on users' trust in automated systems.

In this work we described the existing tools for trust measurement. We selected the trust measurement scale of Jian et al. (2000) for our experiment, and we suggested measurement of the overall trust in systems.

The study presented experimental results about the impact of interface usability on the development of trust. To conduct this experiment usability evaluation of a novel remote sensing application IDFS developed by LM Canada has been performed. Two evaluation methods have been used: heuristic and walkthrough (with expert users). During the evaluation a particular importance was paid to those problems which, according to the proposed guidelines (see Section 3.3), could have a negative impact on trust. The two evaluation methods revealed a total of 46 problems. These included 20 high gravity problems (18 problems found by heuristic evaluation and 14 found by walkthrough, many of which were revealed by both methods), 19 medium priority problems (13 found by heuristic evaluation and 13 found by walkthrough), and seven low priority problems (5 found by heuristic evaluation and 3 found by walkthrough).

We have developed recommendations for improving the interface of IDFS. These recommendations have been implemented by the developers of the IDFS to improve the interface.

The new interface provided enhanced compatibility with the user expectations, a better informative feedback, more visual clarity and more system traceability. Both the new and the old interfaces have been tested by users in the domain of remote sensing. Users' trust was measured. The results of trust measurement with the scale of Jian et al. (2000) revealed that the users' trust in the new interface was increased by a factor of about 2. Thus, the hypothesis that better usability promotes trust was supported by the measurements.

The IDFS is a novel system, so the users did not have any prior experience with it. Hence, trust could not be based on the analytical trust, i.e. trust developed as the result of analyzing the system behavior based on one's experience, knowledge and mental model of the system. In our case the trust could only be due to affective and analogical factors. This is also consistent with the proposition of Miller (2005) that for novel systems the affective and analogical trusts play a key role in trust development. Thus, we can conclude that the enhancement of interface qualities such as compatibility of the system's behavior with the user's expectations, informative feedback, visual clarity and traceability all support the effective and analogical trust development.

Our study showed that the Jian et al.'s (2000) measurement scale is well suited for measuring the initial trust in novel systems. However, the tests also revealed some weaknesses in this scale:

- The scale does not provide definitions for the terminology used. The experimental study revealed that the users had difficulties in understanding the terminology. Even when a particular terminology seemed to be familiar, its perception in the context of the measurement scale was different from one user to another.

- The scale provides 12 answers for 12 trust related items, but it does not provide a single unified result for the overall trust.

To have consistency in the subjects' perception of the items of the scale we suggest to provide explicit explanations for each item of the measurement scale.

We proposed measurement of a single "overall trust" in the system which complements the measurement scale of Jian et al. (2000). First of all it gives a possibility to derive results for both individual trust items and for the overall trust. Because the equations for averaging the measured values for different items and the results for individual users are linear, one can derive the same value for the overall trust using a different order of averaging, i.e. first finding the overall trust for individual users, and only then deriving the average for all users participating in the evaluation.

The results of this study provide bases for future work.

Future research could explore and reveal the differences between the weights of items in the measurement scale of Jian et al. (2000). One of the necessary conditions to do that is to involve a larger number of users to measure trust for different tasks and with more functionalities of the system in order to calculate with sufficient precision the deviation of each individual item from the mean value (overall trust). For example, a larger dispersion could be an indication of a larger confusion in the meaning of item. However, dispersion cannot be a single criterion for assigning the weight. Other factors should be taken into account, such as direct bearing of each item to trust.

This study also investigated the impact of interface usability on the development of trust. However, it does not show the impact of individual usability qualities, such as help and guidance, consistency, error prevention, flexibility and rapidity of use, on trust. One way to measure the impact of different interface qualities on trust is to design an experiment with two versions of interface: one with the interface qualities and the other without them. The comparison of these results would show which interface qualities play major role in trust development.

REFERENCES

ALHAKAMI, A. S. and SLOVIC, P. (1994). A psychological study of the inverse relationship between perceived risk and perceived benefit. *Risk Analysis,* 14(6), pp. 1085-1096.

ASHCRAFT, M.H. (1994). Human Memory and Cognition, 2nd Edition. Harper Collins, Glenview, IL

BISANTZ, A.M., FINGER, R., SEONG, Y. and LLINAS, J. (1999). HumanPerformance and Data Fusion Based Decision Aids. *Proc. of the 2nd International Conference on Information Fusion – Fusion '99,* Vol. 2, Sunnyvale, CA, pp. 918-925.
http://www.eng.buffalo.edu/~bisantz/pubs/funssion99paper.pdf

BISANTZ, A.M. and SEONG, Y. (2001). Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics*, Vol. 28 (2), pp.85-97.

BRIGGS, P., BURFORD, B. and DRACUP, C. (1998). Modeling self-confidence in users of a computer-based system showing unrepresentative design. *International Journal of Human-Computer Studies*, Vol. 49, pp. 717-742.

BOWERS, C. A., OSER, R. A., SALAS, E. and CANNON-BOWERS, J. A. (1996). Team performance in automated systems. *In R. Parasuraman, & M. Mouloua, (Eds.), Automation and Human performance: Theory and Application,* pp. 243-263. Mahwah, NJ: Lawrence Erlbaum Associates.

BURT, R.S. and KNEZ, M. (1996). Trust and third party gossip. *In R.M. Kramer &Tyler(Eds.), Trust in organisations: Frontiers of theory and research,* pp. 68-89. Thousand Oaks, CA: Sage.

CUMMINGS, M. L., MITCHELL, P. M. and SHERIDAN, T. B. (in review). Human supervisory control challenges in Network Centric Operations. *In Human Systems Information Analyses Center (HSIAC) (Ed), State of the Art Report.* Dayton, OH: AFRL.

DIJKSTRA, J.J., LIEBRAND, W.B.G. and TIMMINGA, E. (1998). Persuasiveness of expert systems. *Behaviour and Information Technolog*, vol. 17, pp. 155–163.

DZINDOLET, M., PIERCE, L.G., BECK, H.P. and DAWE, L. (1999). Misuse and disuse of automated aids. *Proc. of the Human Factors an Ergonomics Society 43rd Annual meeting,* pp. 339-343.

DZINDOLET et al. (2000). Building trust in automation. *Paper presented at Human Performance, Situation Awareness and Automation: User-Centered Design for the New Millenium, The 4th Conference on Automation Technology and Human Performance and the 3rd Conference on Situation Awareness in Complex Systems,* October 15-19.

DZINDOLET, M. T., PIERCE, L. G., BECK, H. P. and DAWE, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors,* Vol. 44, pp. 79–94.

DZINDOLET M.T., PETERSONA, S.A., POMRANKY, R.A., PIERCE, L. G. and BECK, H. P. (2003). International Journal of Human-Computer Studies, Vol. 58, issue 6, pp. 697–718.

EATMP Human Resources Team (2003a). Guidelines for Trust in Future ATM Systems: A Literature Review. *HRS/HSP-005-GUI-01. Edition Date: 05.05.2003 Edition 1.0. Released Issue*. Brussels: EUROCONTROL. http://193.221.170.226/humanfactors/gallery/content/public/docs/DELIVERABLES/HF32-HRS-HSP-005-GUI-01withsig.pdf

EATMP Human Resources Team (2003b). Guidelines for Trust in Future ATM Systems: Principles. *HRS/HSP-005-GUI-03. Edition Date: 05.05.2003 Edition 1.0.Released Issue.* Brussels: EUROCONTROL.

http://193.221.170.226/humanfactors/gallery/content/public/docs/DELIVERABLES/HF34-HRS-HSP-005-GUI-03withsig.pdf

EATMP Human Resources Team (2003c). Guidelines for Trust in Future ATM Systems: Measures. *HRS/HSP-005-GUI-01. Edition Date: 05.05.2003 Edition 1.0. Released Issue.* Brussels:EUROCONTROL
http://193.221.170.226/humanfactors/gallery/content/public/docs/DELIVERABLES/HF33-HRS-HSP-005-GUI-02withsig.pdf

ENTIN, E. B., ENTIN, E. E. and SERFATY, D. (1996). Optimizing aided target-recognition performance. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting,* pp. 233–237. Santa Monica, CA: Human Factors and Ergonomics Society.

FOX, J. M. and BOEHM-DAVIS, D. A. (1998). Effects of age and congestion. information accuracy of advanced traveler information on user trust and compliance. *Transportation Research Record,* 1621, pp. 43–49.

FINE, G. A. and HOLYFIELD, L. (1996). Secrecy, trust, nd dangerous leisure: Generating group cohesion in voluntary organisations. *Social Psychology Quarterly,* Vol. 59, pp. 22-38.

GOULD, J.D., BOIES, S.J. and UKELSON, J. (1997). How to design usable systems. *In Helander,M. G., Landauer, T.K., Prabhu, P. (Eds). Handbook of Human-Computer Interaction.* 2nd Edition, Elsevier, North-Holland, pp. 231-254.

GUIDOTTI, P. and TURCHI F. (1995). Knowledge-based systems and distributed processing models, in: J.C. Hage, T.J.M. Bench-Capon, M.J. Cohen, H.J. van den Herik (Eds.), *Legal knowledge based systems JURIX '95: Telecommunication and AI & Law*, Lelystad: Koninklijke Vermande, pp. 43-52, ISBN 90 5458 252 9.

HALL, D. L. and GARGA, A. K. (1999). Pitfalls in DF (and How to Avoid Them), Applied Research Laboratory, the Pennsylvania State University, *EuroFusion99,* 5-7 October.

HOC, J. M. (2000). From human-machine interaction to human-machine cooperation. *Ergonomics*, Vol. 43, pp. 833–843.

HUANG, H., KESER, C., LELAND, J. and SHACHAT, J. (2002). *Trust, the Internet and the digital divide* (RC22511). Yorktown Heights, NY: IBM Research Division.

JIAN, J-Y., BISANTZ, A. M. and DRURY, C. G. (1998). "Toward an Empirically Determined Scale of Trust in Computerized Systems: Distinguishing Concepts and Types of Trust." *In Proceedings of the Human Factors Society 42nd Annual Meeting.* pp. 501-505. Santa Monica, CA: Human Factors and Ergonomics Society. http://www.eng.buffalo.edu/~bisantz/pubs/hfestrust2.pdf

JIAN, J. Y, BISANTZ, A. M. and DRURY, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*, Vol. 4(1), pp. 53 - 71.

JIANG, X., KHASAWNEH, M. T., MASTER, R., BOWLING, S. R., GRAMOPADHYE, A. K., MELLOY, B. J. and GRIMES, L. (2004). "Measurement of Human Trust in a Hybrid Inspection System Based on Signal Detection Theory Measures," *International Journal of Industrial Ergonomics,* Vol. 34(5), pp. 407-419. http://www.ws.binghamton.edu/mkhasawn/Publications/Journal/IJIE%202004%20(Trust).pdf

KIM, J. and MOON,J.Y. (1998). Designing towards emotional usability in customer interface - Trustworthiness of cyber - banking system interfaces. *Interaction with computers,* Vol. 10, pp. 1-29.

ITOH, M., ABE, G. and TANAKA, K. (1999). Trust in and use of automation: Their dependence on occurrence patterns of malfunctions. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics.* Vol. 3, pp. 715–720. Piscataway, NJ: IEEE.

KRAMER, R.M. (1999). Trust and distrust in organisations: Emerging perspectives, enduring questions. *Annual Review of Psychology,* Vol. 50, pp. 569-598.

LEE, J. D. and MORAY, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics,* Vol. 35, pp. 1243-1270.

LEE, J. D. and MORAY, N. (1994). Trust, self-confidence, and operators' adaption to automation. *International Journal of Human-Computer Studies*, Vol. 40, pp. 153-184.

LEE, J. D. and SANQUIST, T. F. (2000). Augmenting the operator function model with cognitive operations: Assessing the cognitive demands of technological innovation in ship navigation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans,* Vol. 30, pp. 273–285.

LEE, J. D. and SEE K. A. (2004). Trust in Automation: Designing for AppropriatReliance. *Human Factors*, Vol. 46, pp. 50-80.
http://www.engineering.uiowa.edu/~csl/publications/pdf/leesee04.pdf

LERCH, F. J., PRIETULA, M. J. and KULIK, C. T. (1997). The Turing effect: The nature of trust in expert system advice. In P. J. Feltovich & K. M. Ford, (Eds.), *Expertise in Context: Human and Machine.* pp. 417-448. Cambridge, MA: The MIT Press.

LEWANDOWSKY, S., MUNDY, M. and TAN, G. (2000). The dynamics of trust: Comparinghumans to automation. *Journal of Experimental Psychology: Applied.* Vol. 6, 2, pp. 104-123.

http://www.psy.uwa.edu.au/Users%20web%20pages/cogscience/Publications_Main.hm

LLINAS, J., BISANTZ, A., DRURY, C. G., SEONG, Y. and JIAN, J. (1998). Studies and analyses of aided adversarial decision-making. Phase 2: Research on Human Trust in Automation. *Center for Multisource Information Fusion,* State University of New York at Buffalo, Technical report . AD-A370937.

http://www.infofusion.buffalo.edu/reports/BISANTZ/reports/AADM%20Pgm%20Trust%20Report%20Main%20Version.pdf

LOEWENSTEIN, G. F., WEBER, E. U., HSEE, C. K. and WELCH, N. (2001). Risk as feelings. *Psychological Bulletin,* Vol. 127, pp. 267-286.

MADHAVAN, P., WIEGMANN, D. A. and LACSON, F.C.(2003). AutomationFailures on Tasks Easily Performed by Operators Undermines Trust in Automated Aids. *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society.* http://www.humanfactors.uiuc.edu/Reports&PapersPDFs/humfac03/madlachf03.pdf

MADHAVAN, P. and WIEGMANN, D. (2005). Cognitive Anchoring on Self-Generated Decisions Reduces Operator Reliance on Automated Diagnostic Aids. *Human Factors,* Vol. 47 (2), pp. 332-341.
http://www.humanfactors.uiuc.edu/Reports&PapersPDFs/humfac05/madwie.pdf

MADHAVAN, P., WIEGMANN, D. A. (2004), A New Look at the Dynamics of Human-Automation Trust: Is Trust in Humans Comparable to Trust in Machines?. *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society.* http://www.humanfactors.uiuc.edu/Reports&PapersPDFs/humfac04/madwieg.pdf

MADSEN, M. and GREGOR, S. (2000). Measuring human-computer trust. In: *Proceedings of Eleventh Australasian Conference on Information Systems*, Brisbane, 6-8 December.

MARSH, S. and MEECH, J. (2000). Trust in design. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems* (The Hague, The Netherlands, April 01 – 06, 2000).CHI'00.ACMPress, New York, NY, pp. 45-46.
http://doi.acm.org/10.1145/633292.633322

MASALONIS, A. J., DULEY, J., GALSTER, S., CASTANO, D., METZGER, U. and PARASURAMAN, R. (1998). Air traffic controller trst in a conflict probe during Free Flight. *Proc. Of the 42nd Annual meeting of the Human Factors and Ergonomics Society.* 1607.

MASALONIS, A. J. and PARASURAMAN R. (1999). Trust as a construct for evaluation of autoated aids: past and present theory and research. *Proceedings Human Factors and Ergonomics Society 43rd Annual Meeting, Santa Monica*, pp. 184-188.

MAYER, R. C., DAVIS, J. H. and SCHOORMAN, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20, pp.* 709–734.

MAYHEW, D.J. (1999). Introduction (chap.1). *In the Usability Engineering Lifecycle.* Morgan Kaufmann, San Francisco, pp. 1-31.

MIL-STD-1472 "Human Engineering Design Criteria for Military Systems, Equipment and Facilities". *Available from the Standardization Documents Order Desk, Building 4D,* 700 Robbins Avenue, Philadelphia, PA 19111-5094.

MILLER, C. A. (2005). Trust in Adaptive Automation: The Role of Etiquette in Tuning Trust via Analogic and Affective Method. In *Proc. of the 1st International Conference on Augmented Cognition*, Las Vegas, NV; July 22-27. http://www.sift.info/English/publications/PDF/Miller-AugCog-Trust-full.pdf

MILLER, C., FUNK, H., GOLDMAN, R., MEISNER, J. and WU, P. (2005). Implications of Adaptive vs. Adaptable UIs on Decision Making: Why "Automated Adaptiveness" is Not Always the Right Answer. In *Proceedings of the 1st International Conference on Augmented Cognition*, Las Vegas, NV; July 22-27. http://www.sift.info/English/publications/PDF/MFGWM-AugCog-DecMak.pdf

MORAY, N., INAGAKI, T. and ITOH, M. (2000). Adaptive automation, trust and selfconfidence in fault management of time-critical tasks; *J. of Experimental Psychol: Applied,* Vol. 6, pp. 44-58.

MOSIER, K. L., SKITKA, L.J., HEERS, S. and BURDICK, M. (1998). Automation bias: Decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology*, Vol. 8, pp. 47-63.

MUIR, B. M. (1987). Trust between humans and machines, and the design of decision aides. *International Journal of Man- Machine Studies,* Vol. 27, pp. 527–539.

MUIR, B.M. (1994). Trust in automation: Part 1. Theoretical issues in study of trust in the study of trust and human intervention in automated systems. *Ergonomics,* Vol. 37 (11), pp. 1905-1922.

MUIR, B. M. and MORAY, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics,* Vol.39, pp. 429-460.

National Transportation Safety Board. (1997). *Marine accident report – Grounding of the Panamanian passenger ship Royal Majesty on Rose and Crown Shoal near Nantucket, Massachusetts,* June 10, 1995 (NTSB/MAR97/01). Washington, DC: Author.

NASS, C.L., FOGG., B. J. and MOON, Y. (1996). Can computers be teammates?*International Journal of Human-Computer Studies*, Vol. 45, 669–678.

NICKERSON, V. J., REILLY R.R. (2004). A model for investigating the effects of machine autonomy on human behavior. *Proceedings of the 37th Hawaii International Conference on System Sciences.*
http://www.stevens-tech.edu/jnickerson/ETSIB01.PDF

NIELSEN, J. (1993). Usability Engineering. *Academis Press, Inc.*

Nuclear Regulatory Commission. (1996). *Human-System Interface Design Review Guideline* (NUREG-0700 Rev. 1, Vol. 1). Washington, DC: U.S. Nuclear Regulatory Commission.

OPPERMAN, R. (1994). *Adaptive user support.* Hillsdale, NJ; Erlbaum.

PARASURAMAN, R. and MILLER, C.A. (2004). Trust and etiquette in high-criticality automated systems. *Communications of the ACM*, Vol. 47 (4), pp. 51-55.
http://archlab.gmu.edu/~rparasur/Documents/ParasuramanACM2004.pdf

PARASURAMAN, R., MOLLOY, R. and SINGH, I. (1993). Performance consequences of automation-induced "complacency". *International Journal of Aviation Psychology,* Vol. 3, pp. 1-23.

PARASURAMAN, R.,MOULOUA, M. and MOLLOY, R. (1996). Affects of adaptive task allocation on monitoring of automated systems. *Human Factors,* Vol. 38, pp. 665-679.

PARASURAMAN, R. and RILEY, V. (1997). Humans and Automation: Use, misuse, disuse, abuse. *Human Factors,* Vol. 39, pp. 230-253. http://archlab.gmu.edu/~rparasur/Documents/ParasRileyHF97.pdf

PARASURAMAN, R., SHERIDAN, T. B. and WICKENS, C.D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems Man and Cyberneics – Part A: Systems and Humans,* Vol. 30, pp. 426-432. http://archlab.gmu.edu/~rparasur/Documents/ParasSherWick2000.pdf

RASMUSSEN, J. (1983). Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models. *IEEE transactions on Systems, Man and Cybenetics,* SMC-13, pp. 257-266.

RASMUSSEN, J., PEJTERSON, A. M. and GOODSTEIN, L. P. (1994). *Cognitive systems engineering.* New York: Wiley.

RAVDEN, S. and JOHNSON, G. (1989). Evaluationg usability of human-computer interfaces: A practical method. *Ellis Hordwood, Chichester.*

REEVES, B. and NASS, C. (1996). The media equation: how people treat computers, television, and the new media like real people and places. *Center for the Study of Language and Information*, Cambridge University Press, Stanford, CA.

REMPEL, J.K., HOLMES, J.G. and ZANNA, M.P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology.* Vol. 49(1), pp. 95-112.

RILEY, V. (1994). Human use of automation. *Unpublished doctoral dissertation, University of Minesota.*

ROBERT, J-M. (2002). Que faut-il savoir sur l'utilisateur pour concevoir des interfaces de qualité. *Dans Boy, G.A. (Ed.). L'Ingenérie cognitive: IHM et Cognition.* Hèrmes, France, 28 pages.

ROBERT, J-M. (1997). Facteur d'utilisabilité des interfaces humain-ordinateur. *Dans Cours IND 6402; Interfaces humain-ordinateur*. Deuxième édition, 2002.

ROBINSON, S.L. (1996). "Trust and Breach of the Psychological Contract," *Administrative Science Quarterly*, Vol. 41, p. 576

ROTTER, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality,* Vol. 35, pp. 651–665.

SHAHBAZIAN, E. (2002). Introduction to DF: Models and Processes, Architectures, Technichues and Applications. *Multisensor Fusion, Proceedings of the NATO Advanced Research Workshop, Multisensor Data Fusion, Pitlochry, Perthshire, Scotland, June 25-July 7, 2000,* @2002 Kluwer Academic Publishers, A.K. Hyder, E. Shahbazian, E. Waltz (Editors), ISBN 1-4020-0722-1

SHERIDAN, T.B. (1988). Trustworthiness of command and control systems. *Proc. Of Analysis, Design and Evaluation of man-Machine Systems 1988*, 3rd IFAC / IFIP / IEA / IFORS Conf., Finland, 14-16 June.

SIMPSON, A. (1992). HCI issues in trust and acceptability. *Defence Evaluation and Research Agency*, Report No. DRA TM(CAD5) 92018, November.

SIMPSON, A. (1995). Seaworthy trust: Confidence in automated data fusion. In: R. Taylor & J. Reising (Eds). *The Human-Electronic Crew: Can we Trust the Team? Proc. of the 3rd Int. Workshop on Human-Computer Teamwork.* Defence Evaluation and Research Agency, Report No. CHS/HS3/TR95001/02, pp. 77-81.

SINGH, I. L., MELLOY, R. and PARASURAMAN, R. (1993). "Automation-induced "complacency": development of the complacency-potential rating scale. *The International Journal of Aviation Psychology*, Vol. 3, pp. 111-122.

SHNEIDERMAN, B (1998). Theories, Principles, and Guidelines (chap.2). *In Designing the user interface.* Eddison-Wesley, Reading, MA, pp. 51-93.

SLOVIC, P., FINUCANE, M. L., PETERS, E. and MACGREGOR, D. G. (2002). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis,* Vol. 24(2), pp. 1-12.

SMITH, L.S. and MOSIER, J.N. (1986) Guidelines for designing user interfaces software. ESD-TR-86-278. United States Air Force, Massachusetts.
http://www.hcibi.org/sam/

SEONG, Y. and BISANTZ, A. M. (2000). Assessment of operator trust in and utilization of automated decision aids under different framing and fault conditions. In Proceedings of the IEA 2000 / HFES 2000 Congress.

SPARACO, P. (1995, January 30). Airbus seeks to keep pilot, new technology in harmony. *Aviation Week and Space Technology,* pp. 62–63.

STEINBERG, A.N., BOWMAN, C. L. and WHITE, F. E. *Revision to the JDL data fusion model,* Third NATO/IRIS Conf., Quebec City, Canada, 1998.
ST. JOHN, M. and MANES, D. I. (2002). Making unreliable automation useful. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting,* pp. 332–336. Santa Monica, CA: Human Factors and Ergonomics Society.

ST. JOHN, M., SMALLMAN, H. S., MANES, D. I., FEHER, B. A. and MORRISON, J. G. (2005). Heuristic automation for decluttering tactical displays. Human Factors, Vol. 47, pp. 509-525.

TAN, G. and LEWANDOWSKY, S. (1996). A comparison of operator trust in humans versus machines. *Proc. of CybErg 1996: The 1st Int. Cyberspace Conf. on Ergonomics. Int. Ergonomics Assoc.*

TAYLOR, R.M., SHADRAKE, R. and HAUGH, J. (1995). Trust and adaptation failure: An experimental study of uncooperation awareness. R. Taylor & J. Reising (Eds), *The Human-Electronic Crew: Can we Trust the Team? Proc. of the 3rd Int. Workshop on Human-Computer Teamwork.* Defence Evaluation and Research Agency, Report No. CHS/HS3/TR95001/02, pp. 93-98.

TSENG, S. and FOGG, B.J. (1999). Credibility and computing technology. *Communications of the ACM,* Vol. 42 (5), pp. 39-44.

WALTZ, E. and LLINAS, J. (1990). Multisemsor Data Fusion. *Artech House Inc.*, Norwood, MA.

WICKENS, C. D. (2003). Aviation displays. In P. Tsang & M. Vidulich (Eds.), *Principles and practices of aviation psychology* (pp. 147-199). Mahwah, NJ: Lawrence Erlbaum Publishers.

WICKENS, C. D., GEMPLER, K. and MORPHEW, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors,* Vol. 2, pp. 99-126.

WICKENS, C. D. and HOLLANDS, J. G. (2000). *Engineering psychology and human performance* (3rd Ed.). Upper Saddle River, NJ: Prentice-Hall.

WICKENS, C. D. and XU, X. (2002). Automation Trust, Reliability and Attention HMI 02, *AHDF Technical Report.*
http://www.humanfactors.uiuc.edu/Reports&PapersPDFs/TechReport/02-14.pdf

WIEGMANN, D. A. (2002). Agreeing with automated diagnostic aids: A study of users' concurrence strategies. *Human Factors,* Vol. 44 (1), pp. 44-50.

YEH, M. and WICKENS, C. D. (2001). Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors*, Vol. 43, pp. 355–365.

ZUBOFF, S. (1988). *In the age of smart machines: The future of work technology and power.* New York: Basic Books.

APPENDIX A

THE JDL MODEL

The JDL model is the most widely used method for categorizing data-fusion related functions. Steinberg et al (1998) define data fusion as "a process of combining data to refine state estimates and predictions". The JDL model was primarily developed for defense applications, but in recent years this technology is applied for other domains such as: remote sensing, medical diagnostics, control systems for large plants (factories, power stations, etc.), air traffic control, communication networks, etc. The detailed description of the JDL DF process model is given by Waltz and Llinas (1990). JDL model has five different levels:

Level 0: Sub-Object Data Assessment: Estimation and prediction of signal/object observable states on the basis of pixel/signal level data association and characterization. This is physical access to raw data. The processes within the Level 0 processing include: Advanced signal processing; Image processing; Synthesis of complex array data to create synthetic information; State estimation.

The output of this processing may be a set of detections/measurements/patterns/features believed to be associated with a single target, or an object detected by one sensor. The detections/measurements may be soft, i.e., sensors provide an m-ary decision, with a measure of uncertainty or confidence, or hard, i.e., the sensors make a binary decision, without a measure of uncertainty or confidence.

Level 1: Object refinement: This process combines location, parametric, and identity information to refine the representations of individual objects. The Level 1 process is often called Multi-Sensor or Multi-Source Data Fusion (MSDF). It deals with object kinematics and identification estimation. Level 0 and 1 processing can be considered sequential of each other.

Level 2: Situation refinement: This process develops a description of the current relationships among objects and events in the context of their environment. Level 2 and

3 processing perform the analysis of the situation, as defined and estimated by Level 1 processing. Level 2 and 3 processes can perform their analyses independently of each other. They also may be activated concurrently in parallel with Level 0 and 1 processes.

Level 3: Threat refinement: Level 3 processing projects the current situation into the future.

Level 4: Process refinement: Level 4 processing may be considered a meta-process, i.e., a process concerned about other processes. This is refinement of the fusion process performance in terms of real-time control and long-term performance as well as the amount and quality of derived or inferred of information.

Figure 2 shows the data flow between the DF levels:

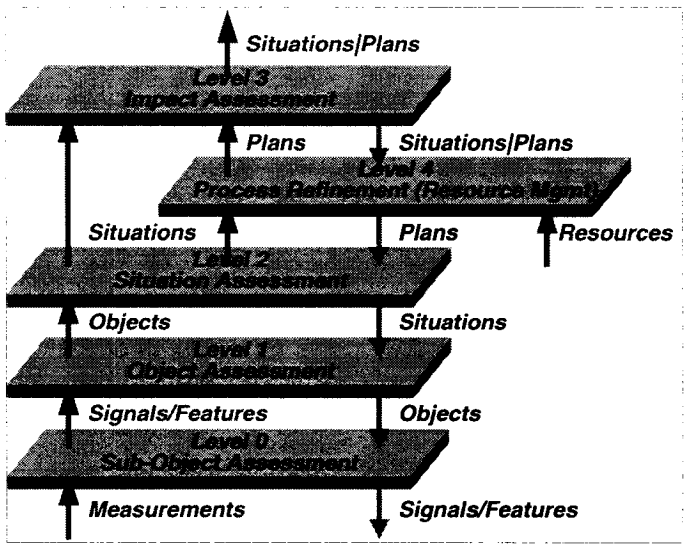


Figure A.1: JDL DF model

Within the JDL model, the DF levels are presented as a sequential flow. In reality there is a lot of parallel/concurrent activation of the functions within a DF system in all levels, and more significantly for levels above 0 and 1 (Shahbazian, 2002).

A good summary of the functionality in each level of processing is provided by Hall and Garga (1999).

APPENDIX B

B.1. HCT RATING SCALE OF MADSEN & GREGOR (2000)

1. Perceived reliability

R1. The system always provides the advice I require to make my decision.

R2. The system performs reliably.

R3. The system responds the same way under the same conditions at different times.

R4. I can rely on the system to function properly.

R5. The system analyzes problems consistently.

2. Perceived technical competence

T1. The system uses appropriate methods to reach decisions.

T2. The system has sound knowledge about this type of problem built into it.

T3. The advice the system produces is as good as that which a highly competent person could produce.

T4. The system correctly uses the information I enter.

T5. The system makes use of all the knowledge and information available to it to produce its solution to the problem.

3. Perceived understandability

U1. I know what will happen the next time I use the system because I understand how it behaves.

U2. I understand how the system will assist me with decisions I have to make.

U3. Although I may not know exactly how the system works, I know how to use it to make decisions about the problem.

U4. It is easy to follow what the system does.

U5. I recognize what I should do to get the advice I need from the system the next time I use it.

4. Faith

F1. I believe advice from the system even when I don't know for certain that it is correct.

F2. When I am uncertain about a decision I believe the system rather than myself.

F3. If I am not sure about a decision, I have faith that the system will provide the best solution.

F4. When the system gives unusual advice I am confident that the advice is correct.

F5. Even if I have no reason to expect the system will be able to solve a difficult problem, I still feel certain that it will.

5. Personal attachment

P1. I would feel a sense of loss if the system was unavailable and I could no longer use it.

P2. I feel a sense of attachment to using the system.

P3. I find the system suitable to my style of decision-making.

P4. I like using the system for decision-making.

P5. I have a personal preference for making decisions with the system.

B.2. CPRS RATING SCALE OF SINGH ET AL. (1993)

Instructions

Read each statement carefully and check one response out of five alternatives in the appropriate box which you feel accurately described your views or experiences. The responses vary on a scale of agreement / disagreement, from "strongly agree" to "strongly disagree". For example:

Statement:

Doing research in a library has been made easier by the introduction of computerized card cataloguing systems.

☐ Strongly agree ☐ Agree ☐ Undecided ☐ Disagree ☐ Strongly disagree

Give your answer for each statement and be sure to place your response in the correct place. Remember, this is an opinion survey and not a test of intelligence or ability. There are no right or wrong answers, only answer that fit your views accurately. Do not skip any question. Time is limited. Do you have any questions?

1. Manually sorting through card catalogues is more reliable than computer-aided searches for finding items in a library.

2. If I need to have a tumor in my body removed, I would choose to undergo computer-aided surgery using laser technology because computerized surgery is more reliable and safer than manual surgery.

3. People save time by using automatic teller machines (ATMs) rather than a bank teller for banking transactions.

4. I do not trust automated devices such as ATMs and computerized airline reservation systems.

5. People who work frequently with automated devices have lower job satisfaction because they feel less involved in their job than those who work manually.

6. I feel safer depositing my money at ATM than with a human teller.

7. I have to tape an important TV program for a class assignment. To ensure that the correct program is recorded, I would use the automatic programming facility on my VCR rather than manually taping.

8. People whose job requires them to work with automated systems are lonelier than people who do not have work with such devices.

9. Automated systems used in modern aircraft, such as the automatic landing system, have made air journeys safer.

10. ATMs provide a safeguard against the inappropriate use of an individual's bank account by dishonest people.

11. Automated devices used in aviation and banking have made work easier for both employees and customers.

12. I often use automated devices.

13. People who work with automated devices have greater job satisfaction because they feel more involved than those who work manually.

14. Automated devices in medicine save time and money in the diagnosis and treatment of disease.

15. Even though the automatic cruise control in my car is set at a speed below the speed limit, I worry when I pass a police radar speed-trap in case the automatic control is not working properly.

16. Bank transactions have become safer with the introduction of computer technology for the transfer of funds.

17. I would rather purchase an item using a computer than have to deal with a sale representative on the phone because my order is more likely to be correct using the computer.

18. Work has become more difficult with the increase of automation in aviation and banking.

19. I do not like to use ATMs because I feel that they are sometimes unreliable.

20. I think that automated devices used in medicine, such as CAT-scans and ultrasound, provide very reliable medical diagnosis.

B.3 SUBJECTIVE RATING SCALE OF LEE AND MORAY (1994)

Table B.1: Trust rating scale of Lee and Moray (1994).

	Questions
1	How high was your self-confidence in controlling the feedstock pump?
2	How much did you trust the automatic controller of the feedstock pump?
3	How high was your self-confidence in controlling the steam pump?
4	How much did you trust the automatic controller of the steam pump?
5	How high was your self confidence in controlling the steam heater
6	How much did you trust the automatic controller of the steam heater?

B.4 . TRUST & AWARENESS SCALE OF TAYLOR ET AL. (1995)

Table B.2: Trust rating scale of Taylor (1995)

Construct	Description
1. Confidence	Confidence in own ability to successfully complete the tasks with the aid of the adaptive automation
2. Self-confidence	Confidence in own ability to successfully complete the tasks
3.Accuracy	Accuracy of own performance on the tasks with the aid of the adaptive automation
4. Self-accuracy	Accuracy of own performance on tasks
5. Automation confidence	Confidence in ability of the machine to support successful completion of the tasks
6. Automation accuracy	Accuracy of machine in successful completion of tasks
7. Automation dependability	The extent to which you can count on the machine to provide the appropriate support to the tasks
8. Automation reliability	The extent to which you can rely on the machine to consistently support the tasks
9. Predictability	The extent to which you can anticipate and expect the machine to support he tasks
10. Predictability	The probability of negative consequences of relying on the machine to support successful completion of the tasks
11. Impact / Survivability	The severity and criticality of adverse or negative consequences of relying on the machine to support successful completion of the tasks

Table B.2: Trust rating scale of Taylor (1995)

12.Decision complexity	The extent to which the machines' decision on when and how to intervene and support the task can be regarded as a simple and obvious choice
13.1 Uncertainty /doubt	The extent to which you have confidence in the machines' decision on when and how to intervene and support he task
14.Judgement / awareness	The extent to which the machines' decision on when and how to intervene and support the task requires assessment, knowledge, and understanding of the task
15- Faith	The extent to which you believe that the machine will be able to intervene and support the tasks in other system states in the future
16. Demand of trust	Level of trust required from you when the machine intervenes and supports the task
17. Supply of trust	Level of trust actually provided by you when the machine intervenes and supports task

B.5 SATI OF EATMP (2003)

SATI Part 1 (*please complete before the start of the day's simulation runs*)

Please tell us who you are and your forthcoming role in the simulation. Thank you
About you:

Name:	
Nationality:	
Sex (M/F):	

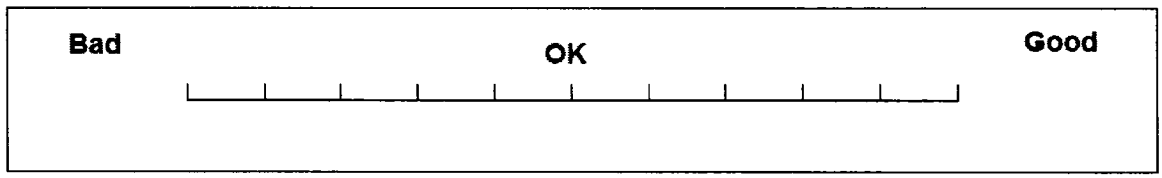
About the simulation:

Date and time:	
Name of simulation project:	
Computer-assistance or automation tools available:	1. 2. 3. 4. 5.
Your simulated sector:	
Your role (planner / executive controller)	

SATI Part 1 (*continued*)

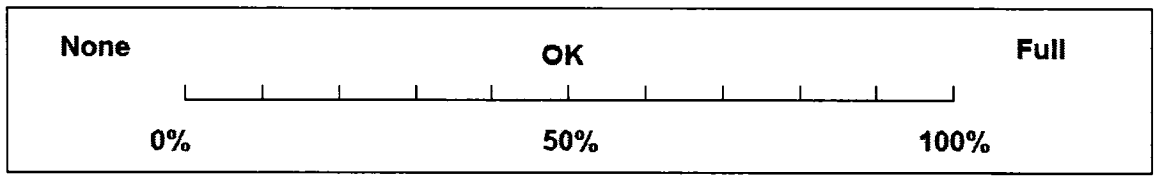
PLEASE COMPLETE AT THE START OF EACH DAY

1. What do you think of the simulation so far? (Please mark the scale with an 'X').



2. Are you prepared to trust the simulated system? Please give your reasons.

3. How much confidence do you have in the simulated system? (Please mark the scale with an
'X').



4. Please give your reasons

SATI Part 2 (*please complete after the end of the simulation runs*)

Please write your name and your last role in the simulation. Thank you.
About you:

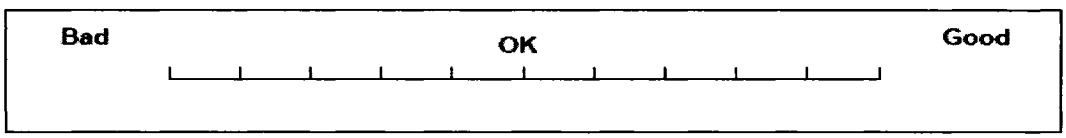
Name:	

About the simulation:

Date and time:	
Name of simulation project:	
Computer-assistance or automation tools available:	1. 2. 3. 4. 5.
Your last simulated sector:	
Your last role (planner / executive controller)	

SATI Part 2 (*continued*)

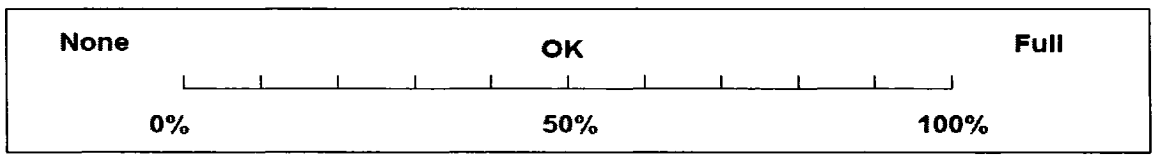
PLEASE COMPLETE AT THE END OF THE DAY'S RUNS
Based on today's runs
1. What did you think of the simulation? (Please mark the scale with an 'X').



2. Were you prepared to trust the simulated system?

No	Yes

3. How much confidence do you have in the simulated system? (Please mark the scale with an 'X').



4. Please give your reasons. If your trust or level of confidence in the system has changed since
the start of the day, please explain why.

SATI Part 2 (*continued*)

PLEASE COMPLETE A SEPARATE SHEET FOR EACH AVAILABLE AUTOMATION TOOL.

5. Please judge each automation tool against the following factors (mark each scale with an 'X').

Name of automation tool:

_____

1. Is the automation tool useful?

Not useful - 5 0 + 5 *Useful*

2. How reliable is it?

Not reliable - 5 0 + 5 *Reliable*

3. How accurately does it work?

Not accurate - 5 0 + 5 *Accurate*

4. Can you understand how it works?

Not understand - 5 0 + 5 *Understand*

5. Do you like using it?

Dislike - 5 0 + 5 *Like*

6. How easy is it to use?

Difficult - 5 0 + 5 *Easy*

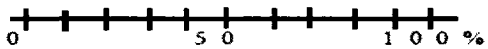
6. Please rank these factors in order of relative importance. Number them from 1 (*least important)* to 6 (*most important*). Please use each number once only.

Name of automation tool:_____	
Usefulness	*ranking:*
Reliability	*ranking:*
Accuracy	*ranking:*
Understanding	*ranking:*
Liking	*ranking:*
Ease of use	*ranking:*

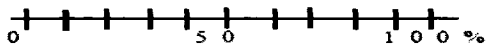
SATI Part 2 (*continued*)

LOOKING BACK OVER THE DAY'S SIMULATION RUNS:
7. Please rate your amount of confidence in each of these five dimensions.
Please mark each scale with an 'X'.

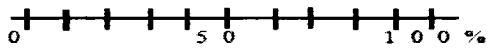
1. Confidence in automation tools



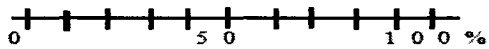
2. Confidence in simulation



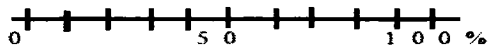
3. Self-confidence



4. Confidence in controller colleagues



5. Confidence in pilots



8. Would you work live traffic with the tools? In your opinion, what changes would the automation need so that your trust and confidence would be increased?
If there are any other factors which influence your trust in an ATC system, or if you have any general comments, please write them here.

Thank you for completing this questionnaire.

APPENDIX C

C.1. SCENARIOS FOR IDFS APPLICATION

Scenario 1

Task: Open the image, create Region of Interest (ROI)

Since no big engineering is possible, all interpretations, all operations will be carried out based on ROI.

Open an image

Click on "File" on the tool bar of the "IDFS" window (Figure 2), from the menu select the "Open HDF5 file" A window with directories and files will appear (Figure C.1).

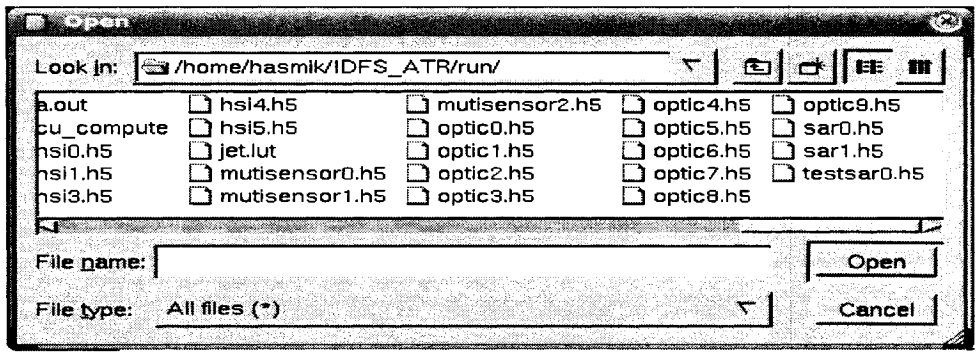


Figure C. 1

Select radar (e.g. sar0.h5) or optical image file (e.g. optic0.h5), click on "Open". Now double-click on "/" on the IDFS window (Figure C.2). Select radar (e.g. sar0.h5) or optical image file (e.g. optic0.h5), click on "Open". Now double-click on "/" on the IDFS window (Figure C.2). The HDF browser for radar images (in the case of radar image) and already created ROI (if available) will appear on the window.

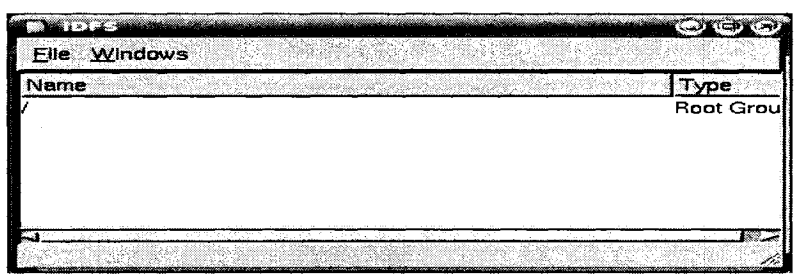


Figure C. 1

Click on "Polarimetric", then on"RAW_HH_AMPL".The radar image will be displayed on the image window
(Figure C.3). To display a particular value of any pixel/intensity on the image, press the middle button of the mouse (Figure C.3).

Figure C. 3

This will display the raw (unprocessed) value of the sensor (2nd line on the figure) and the X, Y coordinates of the pixel (1st line on Figure C.3).
To enhance the image you can right click on image, on the pop-up menu select "Display", then "Linear". You can also select "Histogram Equalization" or get a colored image be selecting "Colormap display".

Create a ROI

You can create any region of interest (ROI) on the main image. Right-click, select "Region" from menu, and then select "Create a region". Click on the corner of the desired region and drug with the mouse. The software permits to create only rectangle shape of region. On the appeared window enter the name of created ROI. After an update of the HDF browser the name of ROI will appear on the browser.

Scenario 2

Task: Compute different features for a selected ROI

In order to carry out conduct the Data Fusion, in the first place you need to compute different features.Open the ROI on HDF browser. Right-click, choose the "Select the

existing region" from "Region" menu. On the appeared window check the needed region (Figure C.4). Now any feature can be executed for selected ROI.

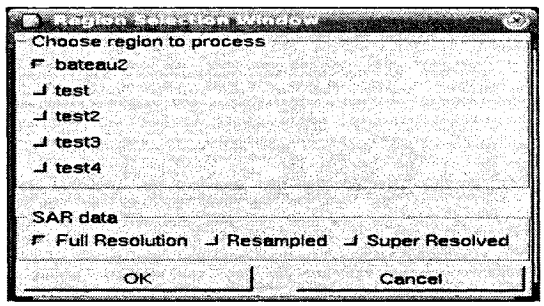


Figure C. 2

For radar images right-click, from the menu "Radar Function"select the feature you need to compute. For hyperspectral images righ select the features from the "Vegetation Index" and for ikonos image select the features from the menu "Ikonos Function".

Features for radar images are the following: Claude's Decomposition, Claude Classifier, Backscatter Coefficient, Scattering Behavior, Amplitude Image, Intensity Image, Equalized Images, Compute Textures, Compute Polarization Response, PWF, Vorticity Index, Odd Even Decomposition, Cameron CTD, SSCM, Polarimetric Discriminators, Subaperture Coherence, Edge extraction, Ratio Line Extraction, CC Line Extraction, Freeman Decomposition, Freeman Approximation, Super Resolution.

Features for Ikonos images (optic images with four bands) are the following: Band Sharpening, Edge Detection, NDVI, GLCM Textures.

Features for hyperspectral images are the following: Compute NDVI, Compute NDWI, Compute SAVI, Compute OSAVI, Compute TCARI, Compute PRI, Compute WBI, Compute MSAVI, Compute ratio TCARI/OSAVI, Estimate Chlorophyll Content.

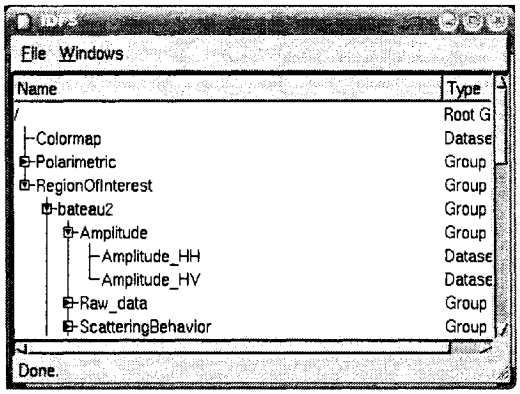


Figure C. 5

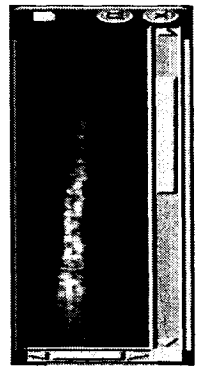


Figure C. 6

After the feature is selected, computed, and HDF browser is updated, it will appear in a hierarchy menu of that ROI on the browser (Figure C.5). Each click on any feature on HDF browser will decline the window of corresponding feature (Figure C.6).

By pressing the middle button of the mouse on the ROI will can receive the coordinates and the reflection value of that corresponding pixel (as for the case of the full radar image).You do not need to select the ROI again in order to extract a new feature for the same ROI. To delete a ROI: right-click on the window "IDFS", select "Dataset & Group", then click on "Delete Dataset". You can not open several windows at the same time.

Scenario 3

Task: Compute the CFAR (target detection) for sea-based dataset

This feature shows in which part of the image exist variations. Right-click, select "CFAR Detection" in menu of "Target Detection". On appeared "CFAR Configuration Window" enter the following parameters (these are for sea based database) (Figure C.7).

CFAR Type: Weibull;	Band: 4
Channel: HH channel	Threshold: 0.05
Object size: 60	Prob. false alarm: 0.01
Window dimension: Width:81	

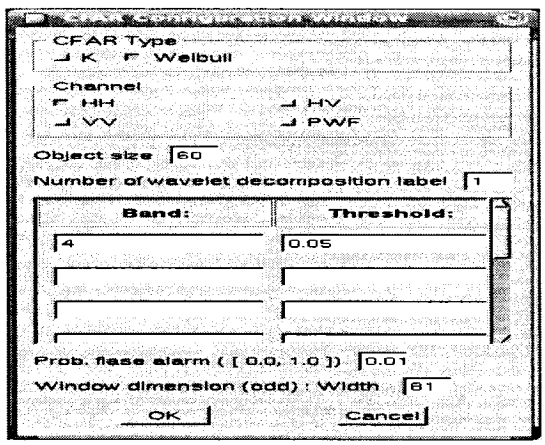


Figure C. 7

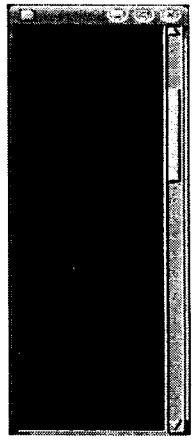


Figure C.8

In HDF browser highlight "Amplitude H"feature. Right click and in "Dataset & Group" menu select "Set as SAR Info". Now open the appeared feature CFAR and you can see the possible target surrounded by yellow boxes (Figure C.8).

Scenario 4

Task: Fusion of Segmentation of several features of the RO

This task should be conducted before the task "Image interpretation". Select the needed ROI, right-click and select the "Region Growing" in the menu "Radar Function". On the appeared window "RegionGrowing Config" (Figure C.9) enter the following parameters and click "OK":

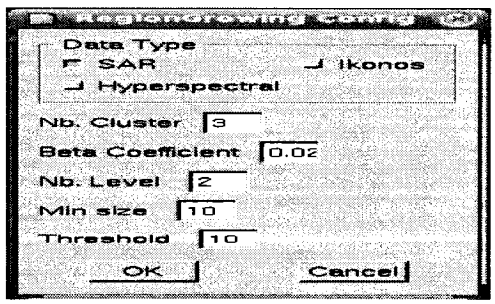


Figure C.9

Data Type: SAR

Nb. Cluster: 2 or 3

Beta Coefficient: 0.02

NbLevel: 2

Min size: 10

Threshold: 10

On the next appeared window select the needed features you find appropriate for the fusion and click "OK" (Figure C.10).

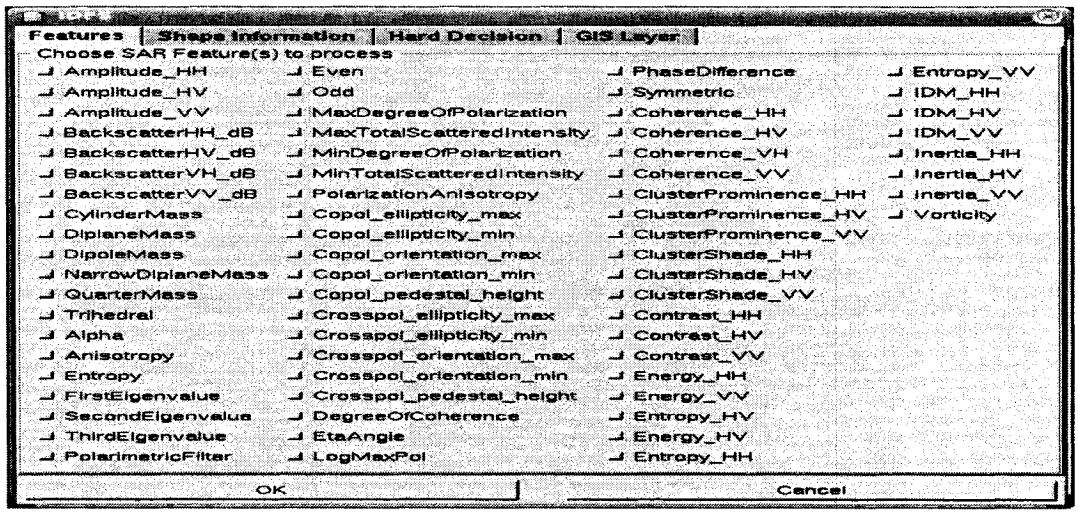


Figure C.10

The system will execute the fusion and reveal the result of segmentation. In menu of HDF browser will appear the "SegmentationSAR" result. It will reveal all levels you have entered. For each level it will reveal the segmentation result (Level_1 or Level_2). It will also reveal three shape parameters for each level: compactness, elongation and rectangularity.

Compactness = Perimeter / Square root of surface

Elongation = max length of the form / width

Rectangularity: More the object has a rectangular form, the more value is close to "1".

For the case of the ship detection this task is not needed (ship is easy to detect). This task is pertinent for the land detection.

Scenario 5

Task: Build a decision tree

This window gives the user the possibility to enter possible propositions that are present in the ROI. You can select from provided default menu, can modify it or define yourself the possible propositions. This can be defined based on the ground truth or your experience. The propositions can be organized in a way of hierarchy, i.e. by building a decision tree.

An example of decision tree can be the following (Figure C.11):

There are 2 possible propositions on the image: Sea and a ship. Each proposition can be refined more (e.g. the ship can be a military or civil ship, etc).

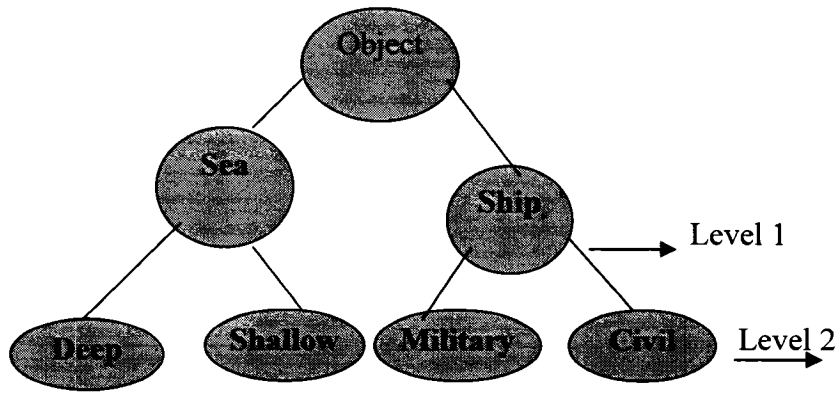


Figure C.11

You can define as much levels you need, e.g. define the civil ship as "Tanker", "Passenger", etc.

Thus, for Level 1 enter the propositions Sea and Ship.

Select the needed ROI. On IDFS tool bar in "Windows" menu choose the "Propositions Window". The Proposition Window has different sections:

- Associated Proposition
- Define Mass Functions Over Current Dataset
- Enter Propositions
- Contextual Relations
- Target Information

- Non-Real Conflict

A click on each tab activates the corresponding section.

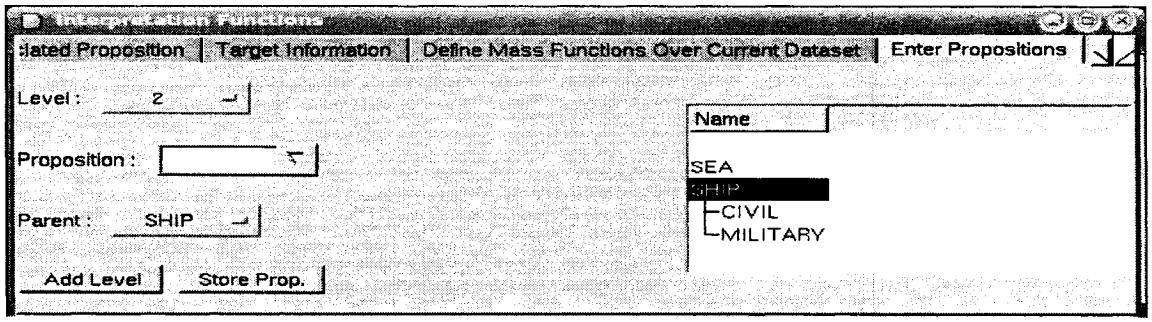


Figure C.12

The window "Define Mass Functions over Current Dataset" provides the default proposition menu. If you want to define another proposition, click on "Enter Proposition".

Enter the needed level, enter the name of proposition in corresponding box and click on "Store Prop.". The names of propositions will appear on the right side of the window (Figure C.12). If you want to enter the next level, click "Add level", select the needed level, select the name of parent from "Parent" menu, enter the proposition name and click on "Store Prop". Now a click on the parent on the right side of the window will decline the created hierarchy. To get this menu in "Define Mass Function over Current Dataset" section, reselect the region once more and this will update the structure.

Scenario 6

Task: Define mass functions for selected feature

Now you need to draw the mass function" for each computed feature. The term mass function can be defined as "the weight of the belief that the certain proposition is present in ROI". You can draw the mass functions by the following methods.

By verifying the reflection values on the image and drawing the corresponding mass functions.

1. By getting the frequencies for each selected proposition and drawing the mass functions according to these frequencies.

Method 1: Defining mass functions by verifying the reflection values.

Select the needed ROI, click on the tab "Define Mass Functions Over Current Dataset" (Figure C.13). Click on any of he buttons of "Propositions supported", the menu with default or defined propositions will appear, select the needed one.

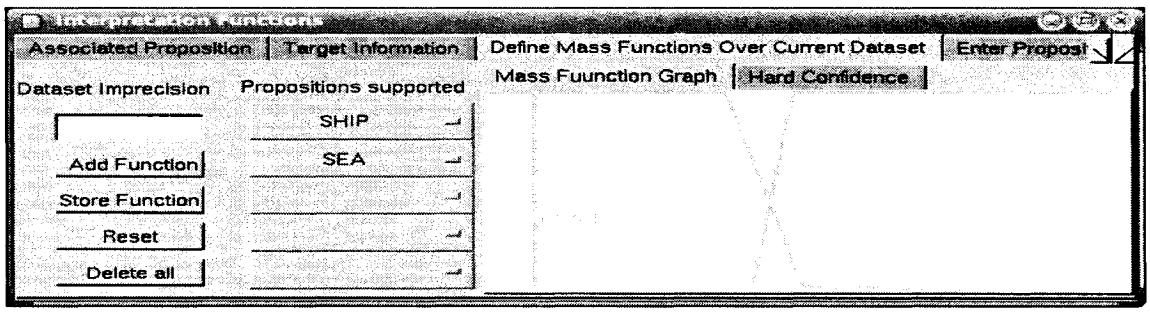


Figure C.13

Select the needed ROI, then from HDF browser select the feature you consider appropriate. On appeared window check the range of reflectance values (by pressed middle button) of the pixels of each proposition (in this case there are only two propositions: Ship and See). The ranges of reflection values on the highlighted feature image and the "Define Mass Functions Over Current Dataset" are always the same.

On the window "Mass Function Editor" click on "Propositions supported", select the needed proposition and on the window draw its mass function.

How to draw mass function:

The mass function for selected proposition should be drawn on the scale of "Mass Function editor". The horizontal axes (Feature value) shows the range of reflectance values for the selected ROI and the vertical axes shows the confidence level (Belief) in reflectance values. Define the pixel value ranges for the proposition (e.g. for the "See", your confidence level that that pixel really belongs to that proposition, and click on corresponding points on the window. Each click will connect the current and the previous nodes. Then click on "Store Function". To draw the mass function for the next proposition click "Add Function" and draw the corresponding mass function. It will be on the same window, all the mass functions will be drawn on the same window (Figure C.13). Imprecision value put 0.01.You draw the mass functions for those features you

consider appropriate for the future fusion process. There are certain features where the result is based on particular thresholding and the mass function is defined by hard decision. In this case you select the proposition and verify its value on the image by pressing the middle button (this will be an integer number). Select that value on value browser in the "Define Mass Functions Over Current Dataset" window (Figure C.14).

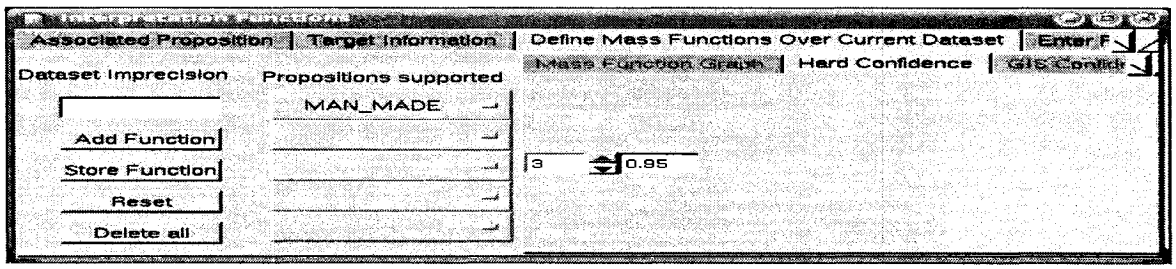


Figure C.14

Enter the belief (e.g. 0.95) in editable part of the window and click the "Store Function".In order to get the fused image of different features: right click and in "Fusion Process" menu select the"For a pixel (fuzzy)". The appeared window will show the features, whose mass functions have been defined.

You can select them all of them or only those you need and click OK (Figure C.15).

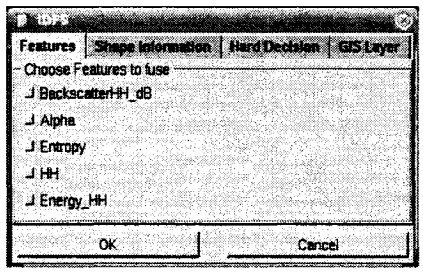


Figure C.15

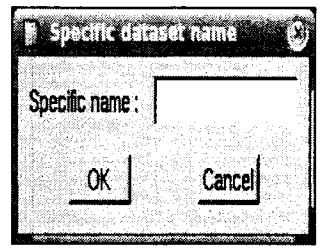


Figure C.16

On the next appeared window (Figure C.16) enter the file name and click OK. Now get the fusion results of belief (BeliefMap) for defined propositions, and for the clusters' visualisation (ClusterMap).To see the belief, open the "BeliefMap" in HDF browser, navigate the image by pressed middle button, and on the image the beliefs and X, Y coordinates will be displayed for each pixel. On Cluster map you can display by selecting the "Colormap Display".

Scenario 7

Task: Define contextual relation between the object in ROI

The section "Contextual Relations" gives the possibility to define the relation between objects in contextual environment. It gives the user the possibility to enter the following:

Object 1 + Adverb + Preposition + Object 2

Adverbs are: Always, Often, Sometimes, Rarely, Never.

Prepositions: Near, Far, Surrounded, Adjacent.

Click on the tab "Contextual Relations" on the window "Interpretation Functions". Define the contextual relation between objects, then press "Store function" button (Figure C.17). For example in the case of propositions "See and Ship" you can enter:

SHIP ALWAYS SURROUNDED by/of SEA

This contextual information is will be used by the system during the fusion process (this will refine the fusion results).

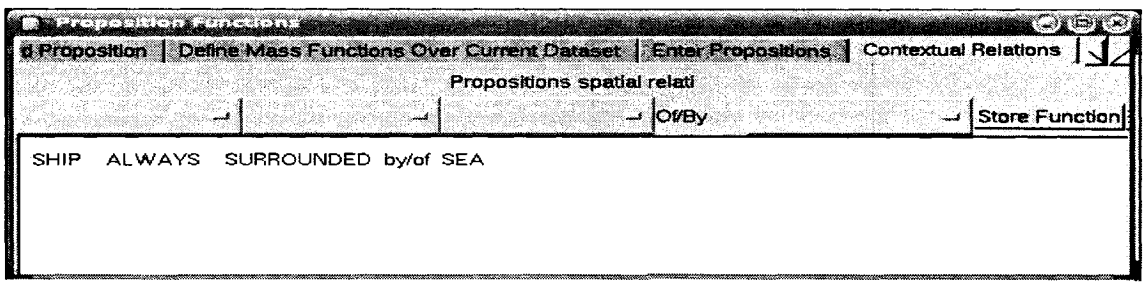


Figure C.17

Scenario 8

Task: Get target information.

This task gives information about the object dimensions. In the first place the feature "Features extraction" should be computed. Right-click, select "Features Extraction" in the menu "Target Recognition". On appeared window (Figure C.18) enter the needed information.

After computation in HDF browser it will appear as "Features". Now open the "Propositions Window" in "IDFS" window and select the section "Target Information". From "Features" select the "ProfileShipImage". Double click on the target on the

displayed image.

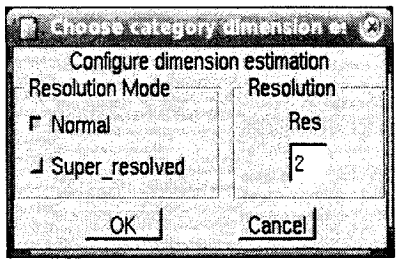


Figure C.18

The dimension information will be displayed on "Target Information" window (Figure C.19). Only the last two dimensions, i.e. "Cartesian Length" and "Angle Length" correspond to real object dimensions (units are in "meters"). This information can be used in DF process.

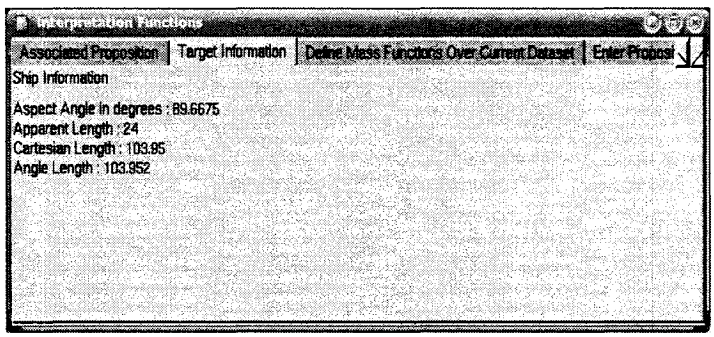


Figure C.19

Scenario 9

Task: Segmentation of one feature of ROI

In order to get the segmentation of one feature click on "Segmentation Window" in the menu "Window" of IDFS main window. In HDF browser highlight the feature you find appropriate and click on the button "Feature". The fingerprints will appear on the window (Figure C.20). With the help of the slider choose the desired number of clusters (it's the number of intersection of the slider with the black zones).Press the button "Normal" and the segmentation process will start. The file with segmentation result will appear on the HDF browser in the menu of feature as "Feature X_Segmentation". To

obtain a colored image of segmentation, from "Display" menu select the "Colormap Display".

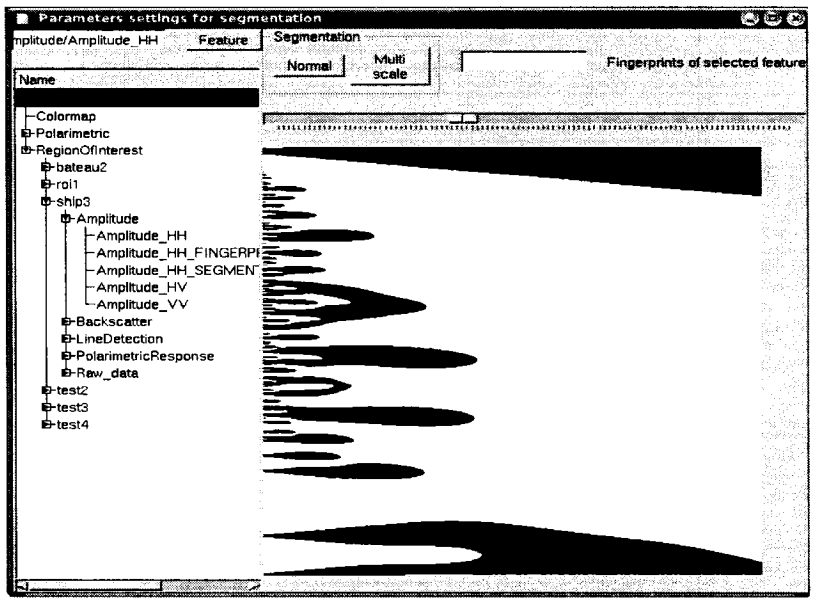


Figure C.20

Scenario 10

Task: To display the 3D-scatter plot (density of scatter plot) using the features available in the ROI

Click on "Windows" in tool-bar of the main window of IDFS, select the "3D plot Window". In the HDF browser on the appeared window select any ROI. There are 3 axes (X, Y, Z) on the window of "3D plotter". For each axis select one feature, i.e. highlight the feature you want to plot, and then push the button of the on corresponding axes. Once all 3 features are entered, push the "Accept" button and the scatter plot will be displayed on the window (Figure C.21). With the help of the 3 sliders (on the left side) you can rotate or zoom the display. Using the other long slider you can browse through density of occurrences (threshold the plot).

Instead of density of scatter plot you can get the cluster of segmentation. It shows the segmentation result, i.e. all the clusters in different colors in 3D plot.

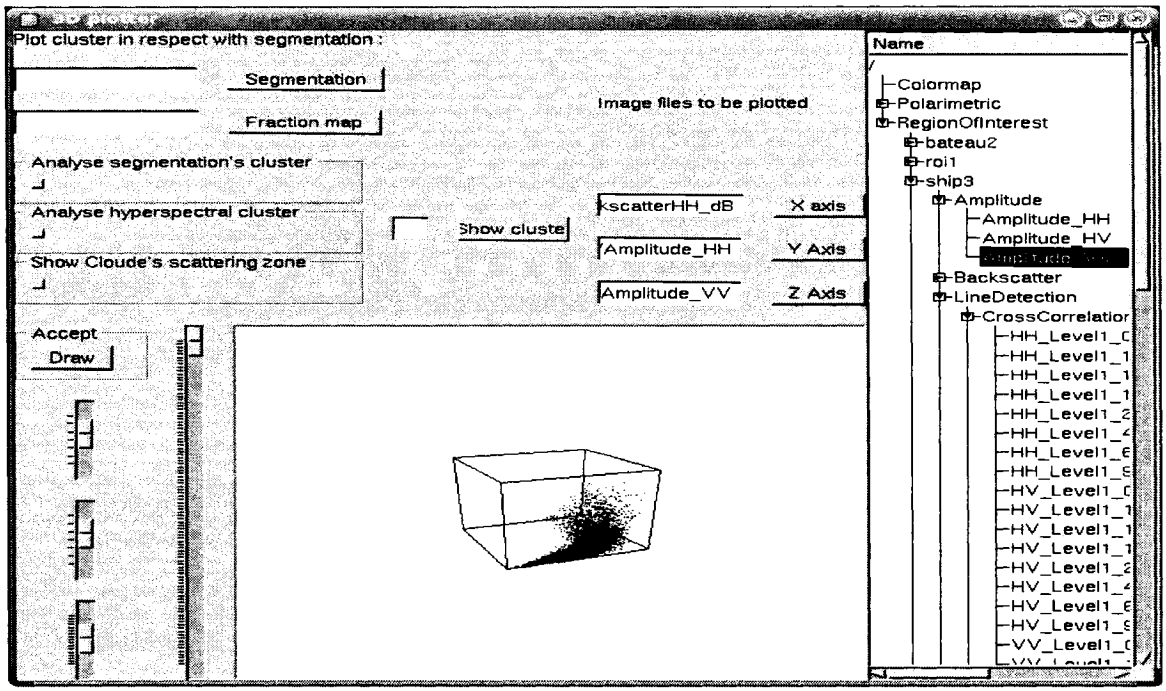


Figure C.21

Select three features, (for X, Y and Z) then highlight and segmentation result (task described in scenario 9) and click on "Segmentation" button. Click on the button "Analyze segmentation's cluster".

The user can also refine this segmentation visualization. The window enables the user to see each cluster separately in 3D plot. Put the segmentation result (that is already computed) in "Segmentation" window. To look at each cluster put each cluster number in "Show Cluster" and click on that button. If the plot will be divided in two or several parts, it means that the cluster can be refined more (e.g. if it's corn field, to get 2 different heights of corn). This window is useful for land cover segmentation and classification.

Scenario 11

Task: Get the Polarimetric Window

Click on "Windows" on the tool-bar, open the "Polarimetric Window", select the ROI, double-click on any location on ROI, and the polarimetric image will appear on the window (Figure C.22).

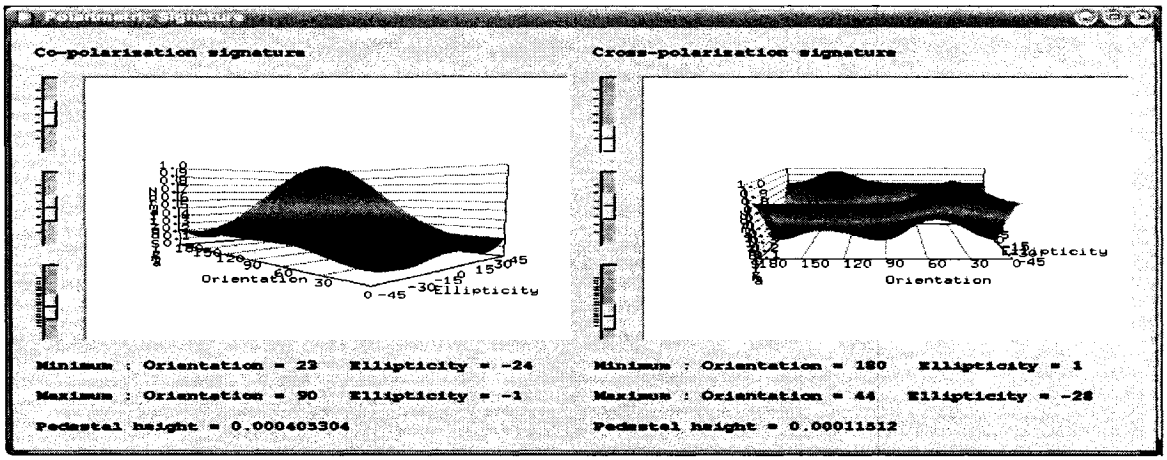


Figure C.22

Scenario 12

Task: Open optic image, create fraction maps, select ROI, navigate through the bands.

Open the hyperspectral image (e.g. "hsi0.h5" in Fig.1). Click on the "File" in IDFS tool bar, then click on "Load Fraction Maps". On the appeared window (Figure C.23) enter

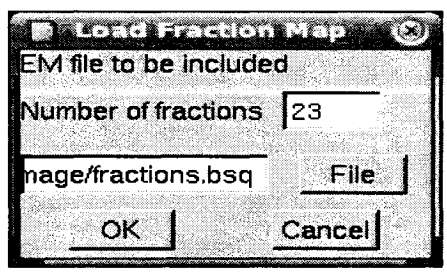


Figure C.23

the number of fractions you find appropriate (for this image enter "23") and insert the file from HSI image (the file with extension .bsq). Now in HDF you will have the defined numbers of fraction maps. Figure C.23. Each Fraction Map represents the fraction (dominance) of particular material present in pixels over the map after spectral unmixing (comparing) with "pure spectra".Select a ROI as in case of radar image. Now you can navigate through the bands of datacube. Click on "Window" in IDFS tool bar,

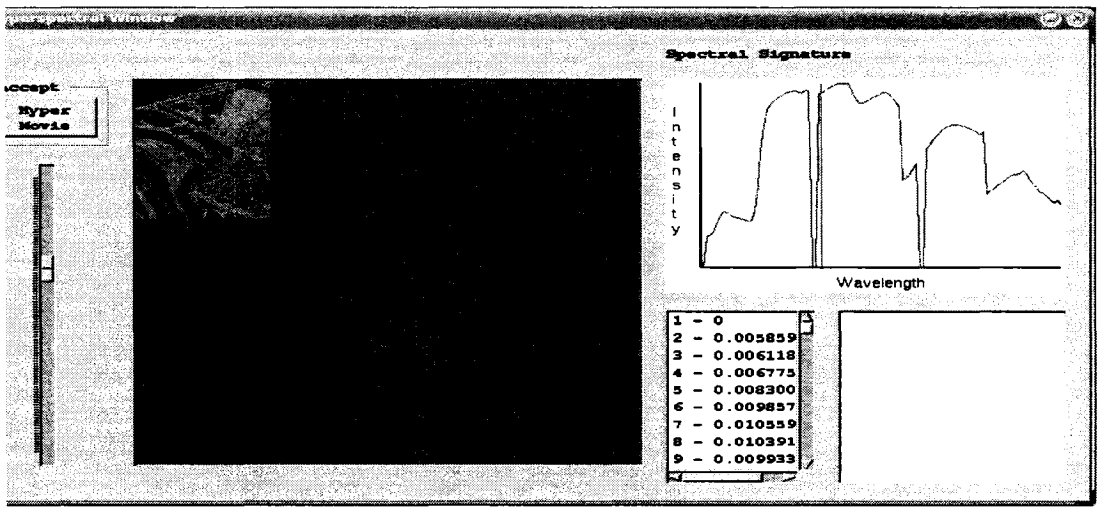


Figure C.24

choose "Hyperspectral Window". Select ROI, and on "Hyperspectral Window" click on "Hyper Movie" button (Figure C.24). By moving the slider you can navigate through the bands of image. In a separate window will appear the reflectance values for all the bands. In a separate window will appear the reflectance values for all the bands. After double-click on any pixel on the image the spectral signature of that pixel will appear in the signature box.

Scenario 13

Task: Compute the features for hyperspectral image, conduct segmentation for different features.

Right click, select the "Hyperspectral Index", and compute the needed features. In HDF browser the features will appear in a directory called "Vegetation Index". Select "Region Growing" from "Hyperspectral Index" menu, enter the needed number of clusters (for

this case: "23"), and all other parameters as in the case for radar image (see Scenario 4) and execute the segmentation. In IDFS window in HDF browser open the "SegmentationHyper" directory. On the appeared window right click, select "Display", then "Colormap display" and you can see the segmentation in different colors. With the help of pressed middle button you can see number of segments.

Scenario 14

Task: Define the decision tree and the mass functions.

Open any feature you find appropriate. Open Proposition Window, enter the propositions. Open "Define Mass Functions over Current Dataset". Since the hyperspectral image contains several propositions, it is more convenient to define mass functions according to the frequencies (second method from Scenario 7). The software will give the frequencies for each selected proposition, so the user can draw the mass functions according to those frequencies. For the second case: select the region, open the image of any feature (e.g. the feature NDVI), and open the "Proposition Window". Enter the propositions in the following hierarchy:

Level 1: Bare_Soil; Vegetation.

Level 2: Parent: Vegetation -> Broadleaf Vegetation; Emerging_Vegetation; Grain_Crop, Grass.

Parent: Bare_Soil -> Bare_Dry_Soil; Bare_Wet_Soil; Road.

Right click on the main window of IDFS, select "Show Frequencies" from "Region" menu. On "Proposition Window" click on the tab "Define Mass functions Over Current Dataset", select any endmember, for which you know the ground truth.

Now on the image select a subregion based on your knowledge of ground truth. The graph of frequencies for that subregion will appear on the window of mass function. Draw the mass function taking into account the graph of the frequencies and also your confidence level as well and click on "Store Function" (Figure C.25).

To draw mass function of a new proposition click on "Add Function" and draw the the mass function for that new proposition. All the mass functions will be on the same window. If the option "Show frequencies" is selected the pressed middle button of the

mouse does not show the reflection values on the image. To get these values you need to select "Stop Showing Frequencies" in "Region" menu.

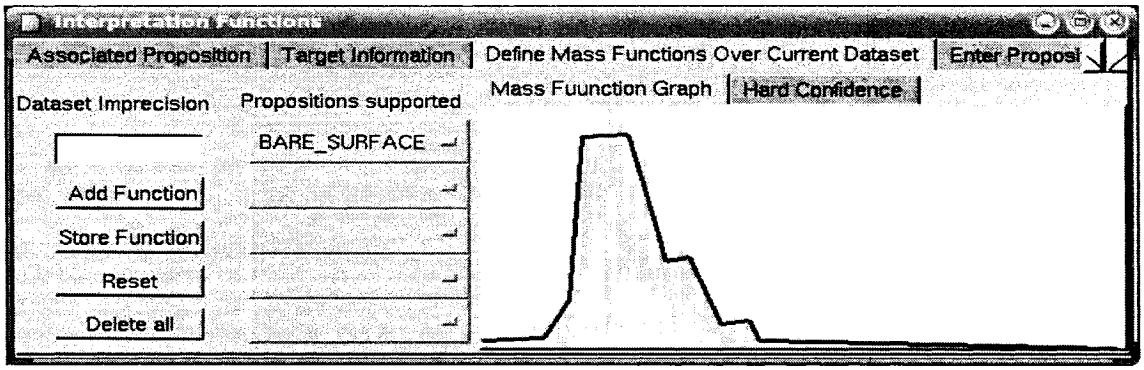


Figure C.25

Scenario 15

Task: For each pixel to get the probability (belief) of presence of endmembers.

Select the ROI, open the "Proposition Window", and click on the tab "Associated Propositions", open one of the computed features (e.g. NDVI). Every double click on the image will trigger the probability of presence of endmembers, in particular the first three endmembers in descending order (Figure C.26).

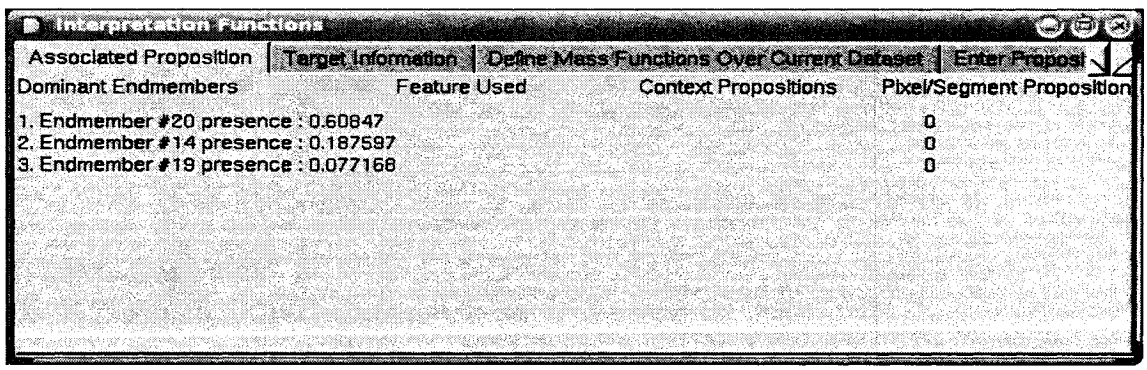


Figure C.26

Scenario 16

Task: To get the fusion result for the presence of defined propositions for each pixel.

Select ROI, right click and select "Start fusing information" from "Fusion Process". Now any double click on any pixel of ROI will decline the window with list of features that have already defined mass functions (Figure C.27).

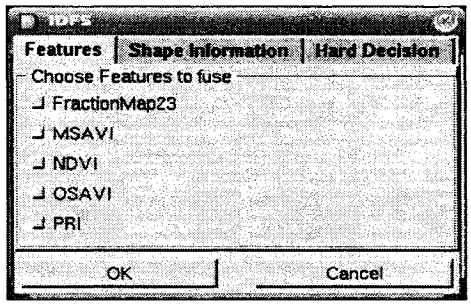


Figure C.27

Select the features you need, click OK. The DF result with the confidence values of the particular pixel will appear on the window "Associated Proposition" (Figure C.28).

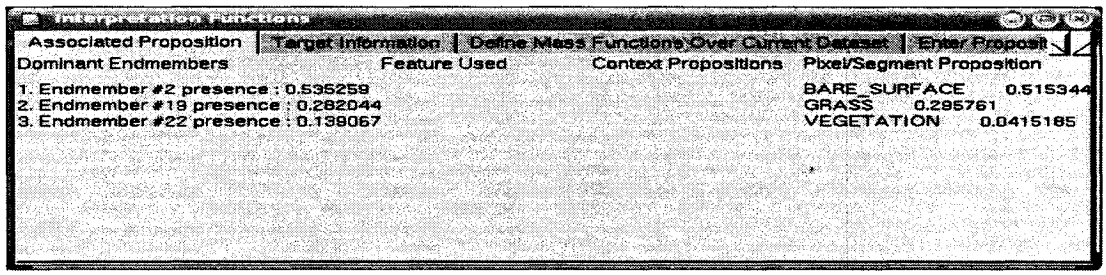


Figure C.28

Scenario 17

Task: To get the fusion result for different features for the whole ROI.

Select ROI, right click, select "For a pixel (fuzzy)" in the menu "Fusion Process". On the appeared window (Figure C.27) select the needed features, click OK. The DF result will appear in HDF browser. Passing with pressed middle button through "BeliefMap"

you will see the belief (probability) for each pixel. The cluster map will give the clusters for the propositions for drawn mass functions (here you can apply the "Colormap" option).

Scenario 18

Task: Define the endmembers based on ground truth

Select the ROI. On the tool-bar select the "Endmember Identification Window"from the menu "Window". Pass through the fraction maps. Based on ground truth and your). For experience enter the endmember name for each fraction map (Figure C.29)

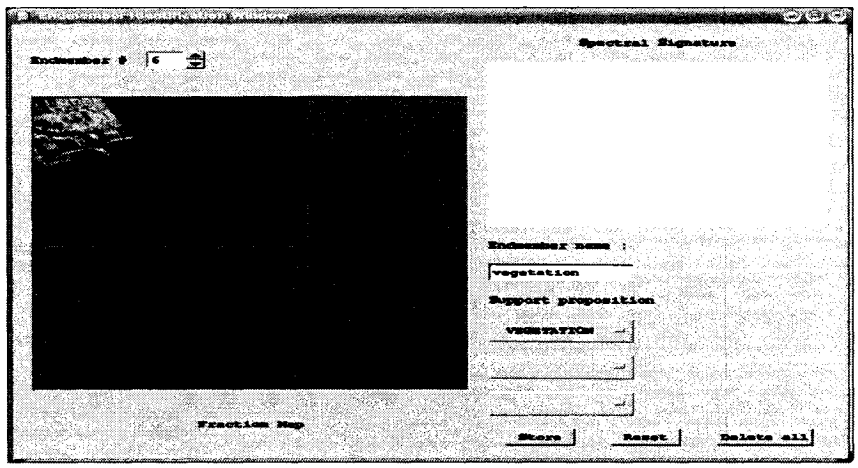


Figure C.29

For each fraction map from "Support proposition" menu list select the proposition you need, then again enter that name in " "Endmember name". Now each double-click on the image will show the endmember names while presenting the information about probability of the presence of the first three endmembers in descending order.

Scenario 19

Task: Create a ROI for Ikonos and SAR images and extract features

Open the "Panchromatic" of Ikonos image. Right-click, choose "Select Independent Region" from "Region" menu. The Ikonos image will open and automatically switch to

a SAR image. This replacement happens because the system checks that both Ikonos and SAR images are available. On SAR image select the ROI and enter the name. After you click on "Accept" button, the Ikonos image will appear on the window. Select the same region as for SAR. The name of the region will be the same, thus no name entry is necessary. On the HDF browser will appear the name of new created ROI.

Extract features for Ikonos and SAR images: For Ikonos features first of all compute the "Band Sharpening", then the feature NDVI. For SAR image in the first place extract the feature "Backscatter", then "Cloud's Decomposition" and "Cloud Classifier". For this particular image these features are sufficient, but you can compute other features as well.

Define the decision tree: Define the decision tree (as described in Scenario 5) in the following hierarchy:

Level 1: Vegetation; Bare_Surface.

Level 2: Parent: Vegetation -> Forest; Grass.

Parent: Bare_Surface -> Man_Made; Water; Bare_Soil; Road; Parking.

Level 3: Parent: Man_Made -> Vehicle; Building.

Scenario 20

Task: Get segmentation for different features and define mass functions.

Right click, select "Region Growing"in "Ikonos Function" menu. On appeared window select the "Panchromatic", Band 1 –> Band 4 (select Bands in PS_MS directory, which is the 2nd set of the Band 1 -> Band 4, Figure C.30) and click on "OK".

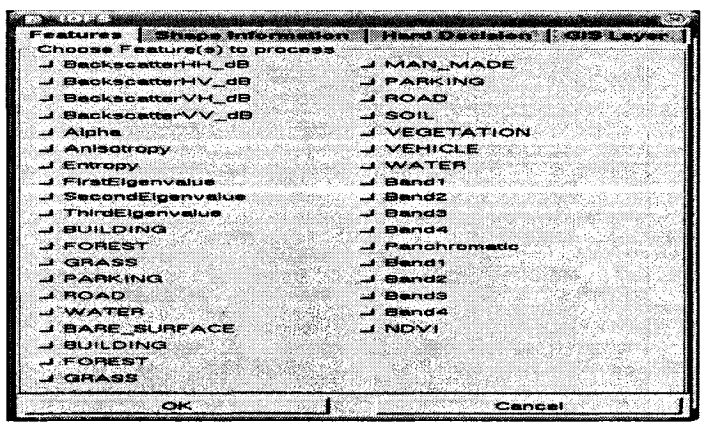


Figure C.30

On the next appeared window ("Region Growing Config") enter the following data:

Right click, select "Region Growing"in "Ikonos Function" menu. On appeared window select the "Panchromatic", Band 1 –> Band 4 (select Bands in PS_MS directory, which is the 2nd set of the Band 1 -> Band 4, Figure C.30) and click on "OK".

On the next appeared window ("Region Growing Config") enter the following data:

Data Type: "Ikonos"	Nb. Level: "1 or 2"
Nb. Cluster: "22"	Min size: "10"
Beta Coefficient: "0.02"	Threshold: "15"

Click "OK", and after the segmentation result will be carried out by the system, it will appear on the HDF browser as "SegmentationIkonos". It will give the segmentation result and also three shape parameters: compactness, elongation and rectangularity.

The same procedure you can apply for the SAR image. For SAR you need to select SAR features and put the number of clusters "12 or 13". In HDF browser it will appear as "SegmentationSAR". Now you can see the beliefs for the propositions.

In order to do that: Open the "Proposition Windows". The first column shows the values for the defined features, the second column shows the beliefs for the context and the final results for the fusion of the features, shape parameters and the context are shown in the last column (Figure C.31)

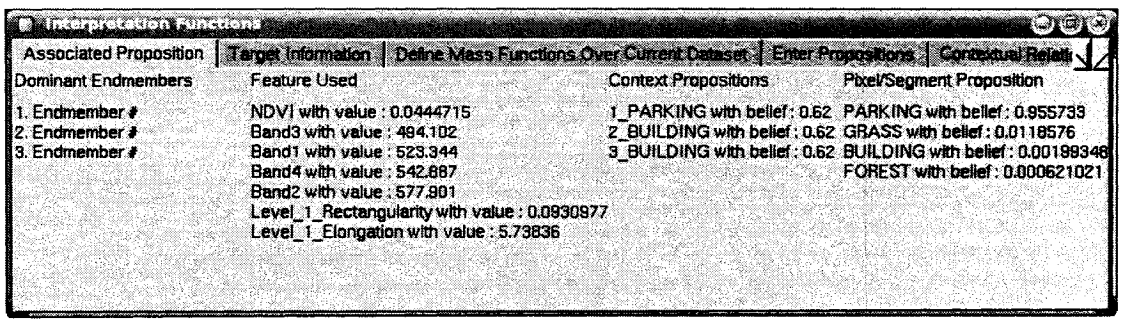


Figure C.31

Define mass functions:

You can define mass functions by the three following ways:

- By definition of pixel values for each proposition (as described in Scenario 6).
- By defining the frequencies for each proposition (as described in Scenario 14).

- By importing the mass functions (if the system has a ready configuration for the defined propositions).

Apply the third method, since the system has the configuration for the propositions of the Ikonos/SAR image in use. In order to import the mass functions in HDF browser highlight the corresponding feature (e.g. PS_MS/Band 1), right-click, from "Dataset & Group" menu select the "Import Mass Functions". Repeat the same for all the features: Band 2, Band 3, Band 4, NDVI. Import also the mass functions for shape parameters "Level_1_Elongation" and "Level_1_Rectangularity" from "SegmentationIkonos" directory of HDF browser.

Apply the same procedure for the SAR image. Import mass functions for Backscattering and H_Alpha_Anisotropy (from the computed feature "Cloud Classifier").

Scenario 21

Task: Get the results of image interpretation

After the segmentation for different features is carried out and the mass functions are defined, you can start the task "Image Interpretation". Right click, select the "Interpret Image" from "Scene Interpretation" menu. On the appeared window (Figure C.32) check the "SegmentationIkonos/Level_X". Now click on "Shape Information" tab. The window will reveal the shape parameters for which the mass functions have been defined (Figure C.33)

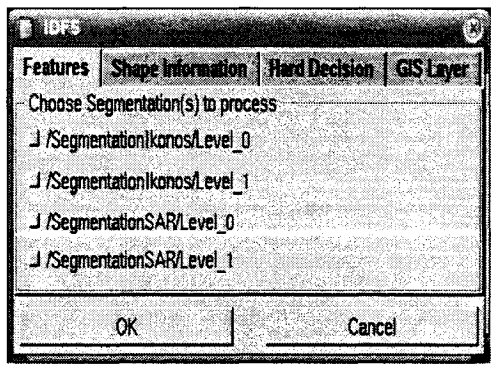


Figure C.32

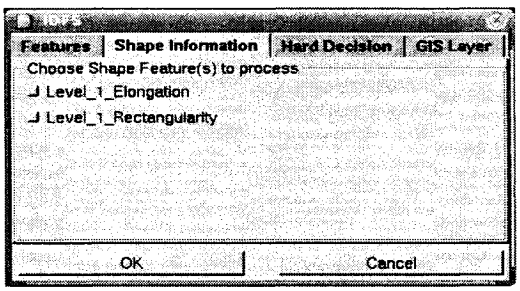


Figure C.33

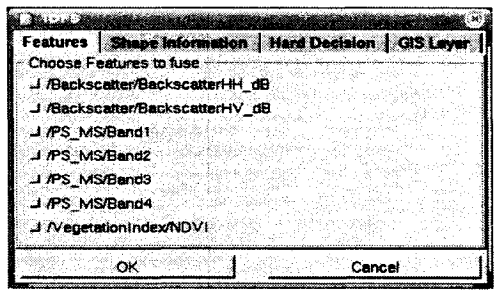


Figure C.34

Check the parameters and click "OK". The next window will appear which will reveal the features with already defined mass functions (Figure C.34). Check the features for Ikonos image and click "OK". Now on HDF browser will appear the results of image interpretation. It will be presented in 2 ways: "DistanceMapsIkonos" and "FeatureMapsIkonos". The first one shows the distance in meters from the nearest proposition. For example, open any proposition image from the "DistanceMapsIkonos" directory (e.g. "Building"), press the middle button and navigate through the image. The location of the proposition itself (e.g. building) shows the value "0". While navigating with pressed middle button, you can see the distance between that mouse position and the selected proposition (in this case the building). The colored image shows better the building placement (Figure C.35).

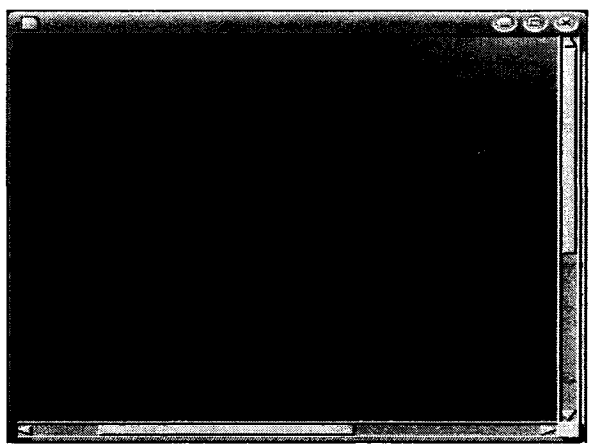


Figure C.35

The "FeatureMapsIkonos" shows the presence of highlighted proposition on the image. The parts that are presented in white color show the presence of the proposition (Figure C.36).

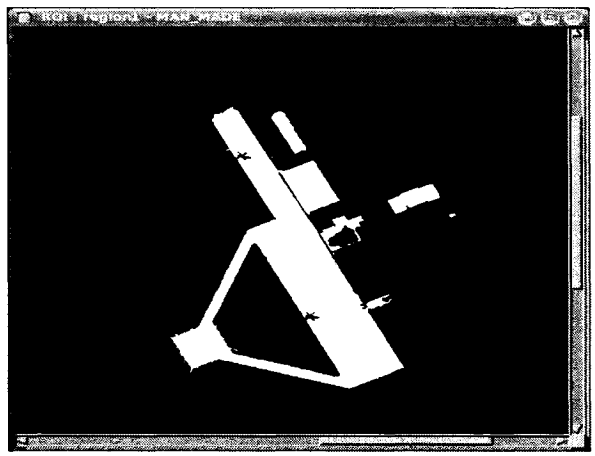


Figure C.36

Apply the same procedure for the SAR image. Select the segmentation level, check the mass functions. Since the mass function for "H_Alpha_Anisotrop" is a hard decision, check its mass function in the section "Hard Decision".

Scenario 22

Task: Fuse the multisensor interpretation

Select the ROI, for co-registration right-click, select the "Select Landmarks" from "Region" menu. On the appeared window (Figure. C.37) check the "SAR -> Ikonos".

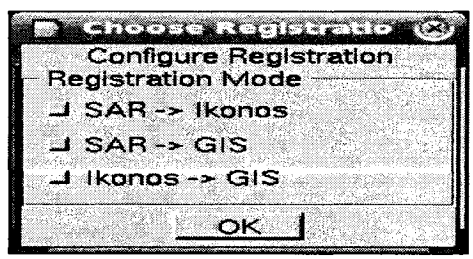


Figure C.37

The Ikonos Panchromatic image will appear. On the image select a point (preferable an angle point, since it's easier to find the same point location on SAR image). Now the

SAR image will appear. Select the same point location as for Ikonos on the SAR image. Select five or six points. Note that the Ikonos and SAR images are in reversed form. To save the landmarks from "Region" menu select "Save Landmarks". To see the results of co-registration, from "Region" menu select the "Landmark Warping". On the appeared window check the "SAR -> Ikonos" option. On the HDF browser this will appear as directories "RegisteredImagesIkonos" and "RegisteredImagesSAR". You can see the results displayed in RGB. In order to do that right-click, following:

- R: TestImageSAR_IKONOS (from the folder "RegisteredImagesIkonos")
- G: Panchromatic (from the folder "Ikonos")
- B: Panchromatic (from the folder "Ikonos")

For SAR image put the following features:

- R: Backscatter HH
- G: Backscatter HV
- B: Backscatter

Now set the registration pointer object by selecting "Set Registration" from "Region". To perform the registration for SAR interpretation select the "Register SAR Interpretation" from "Region" menu and it will reveal you the SAR interpretation. On HDF browser it will appear as folder "FinalInterpretationSAR". Do the same for Ikonos interpretation by selection "Register Ikonos Interpretation".In HDF browser it will appear as "FinalInterpretationIkonos". In HDF HDF browser it will appear as "FinalInterpretationIkonos". Now you can get the beliefs for "FinalInterpretationSAR" or "FinalInterpretationIkonos" for each pixel. In order to do that open that image, then open the section "Associated Propositions" of "Proposition Windows", right-click and select "Set as Segmentation" from menu "Dataset & Group ". Any click on any pixel will reveal the beliefs on the "Proposition Window" (Figure C.38).

To perform the multisensor fusion from "Scene interpretation" menu select "Fuse MS Interpretation". On appeared window check the 2 interpretations:

"/SegmentationIkonos/Level_1" and "/SegmentationSAR/Level_1"(Figure C.39) and click "OK".

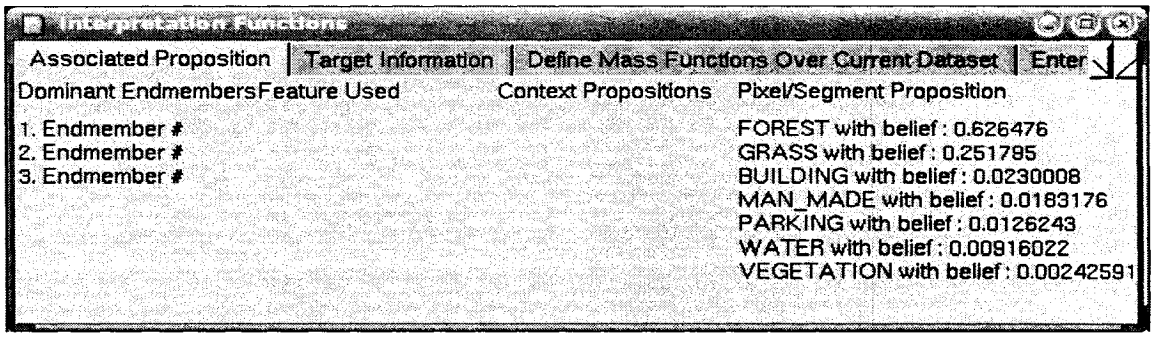


Figure C.38

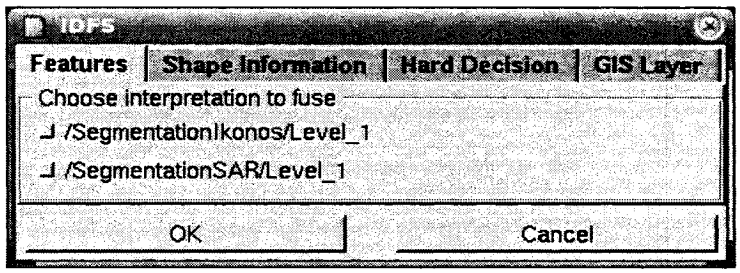


Figure C.39

On the HDF browser you will see all the fusion results. To see the beliefs for propositions (i.e. the weight of presence of that certain proposition) open "Final InterppretationConfidence or "FinalInterpretationSARconfidence", and any double-click on any pixel will reveal "Associated Propositions" of "Proposition Windows" the beliefs for different propositions for that pixel. You can see the beliefs for each pixel (without revealing the propositions) by pressing the middle button and navigating through the image.

Scenario 23

Task: Create your own colors for propositions.

You can create your own colors for each proposition and get the segmentation results according to your defined colors. On "Interpretation Functions" window in the section "Enter Propositions" highlight each needed proposition and right-click next to it in the

column "Color". On the appeared "Select Color" window choose the color you prefer for that proposition and click OK (Figure C.40).

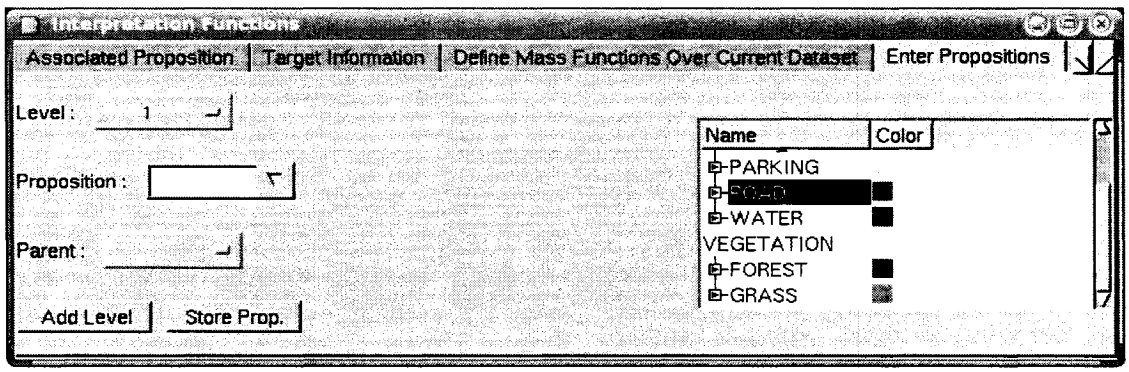


Figure C.40

Open the result of segmentation ("Level_XXX" of the "SegmentationXXX"), right-click, from "Display" menu select the "Land Use Color".

C.2. DEFINE MASS FUNCTIONS (FOR NEW INTERFACE VERSION)

The mass function for selected proposition should be drawn on the scale of "Mass Function editor". The horizontal axes (Feature value) shows the range of reflectance values for the selected ROI and the vertical axes shows the confidence level (Belief) in reflectance values.

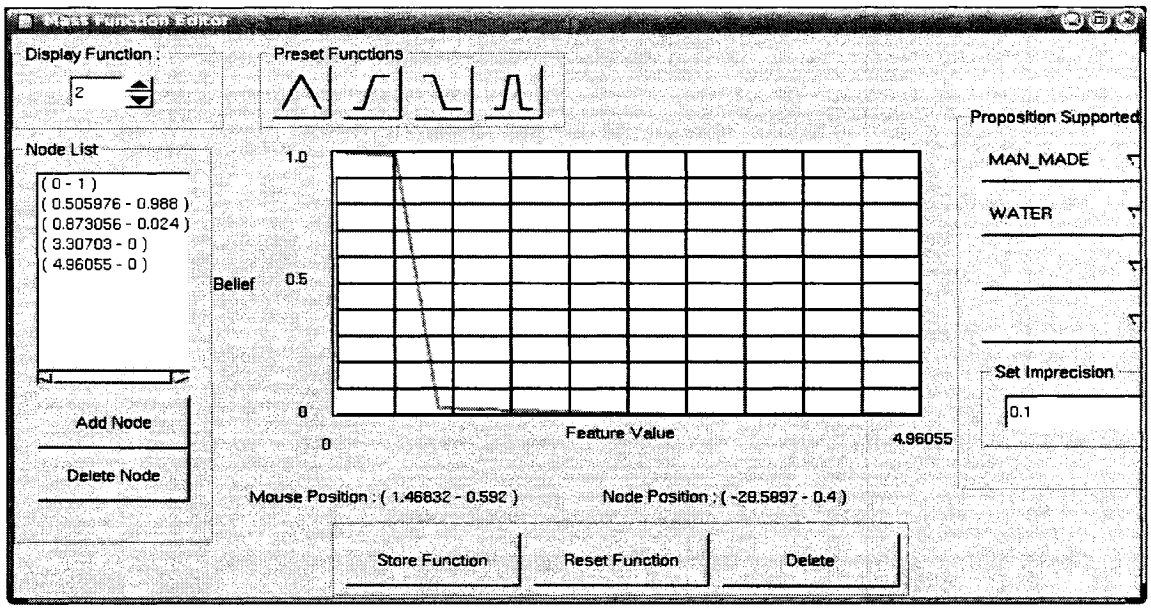


Figure C.40

Pass the mouse over the scale, and below on the window the "Mouse position" will show the range of values for each position of the mouse. Based on defined reflectance values for the selected proposition and the confidence level in the presence of that proposition, click on the corresponding positions on the scale (each click will connect the current and the previous nodes) (Figure 1).

You can draw mass functions by entering each node separately, and you can also draw it by selecting one of the "Preset Functions" and modifying the positions of its nodes by "Add Node" or "Delete Node". In order to delete any node, highlight it in "Node List" and click on "Delete Node". In order to add a node, click the needed point on the scale and click on "Add Node". To store the drawn mass function, click on "Store Function". The mass function for each proposition will be drawn on a separate window. If you need

to see all the defined mass functions, click on “Display Function” on “Mass Function Editor”. To clear the screen but still keep the defined mass function, click on “Reset” button. In order to delete the mass function, click on “Delete” button.

As "Imprecision" value put: 0.1.

The same procedure can be repeated for each proposition. Draw the mass functions for those features, which you consider appropriate for the fusion.

APPENDIX D

GOAL DRIVEN TASK HIERARCHIES OF IDFS APPLICATION

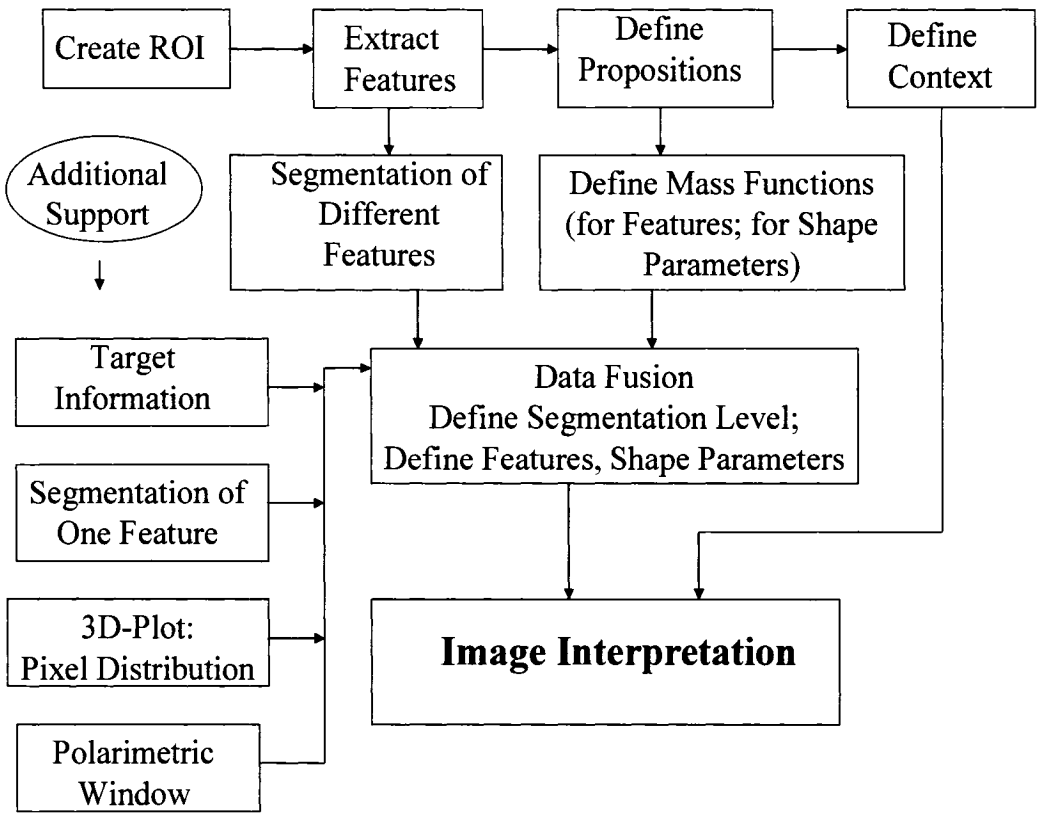


Figure D-1: Task hierarchy for radar image interpretation

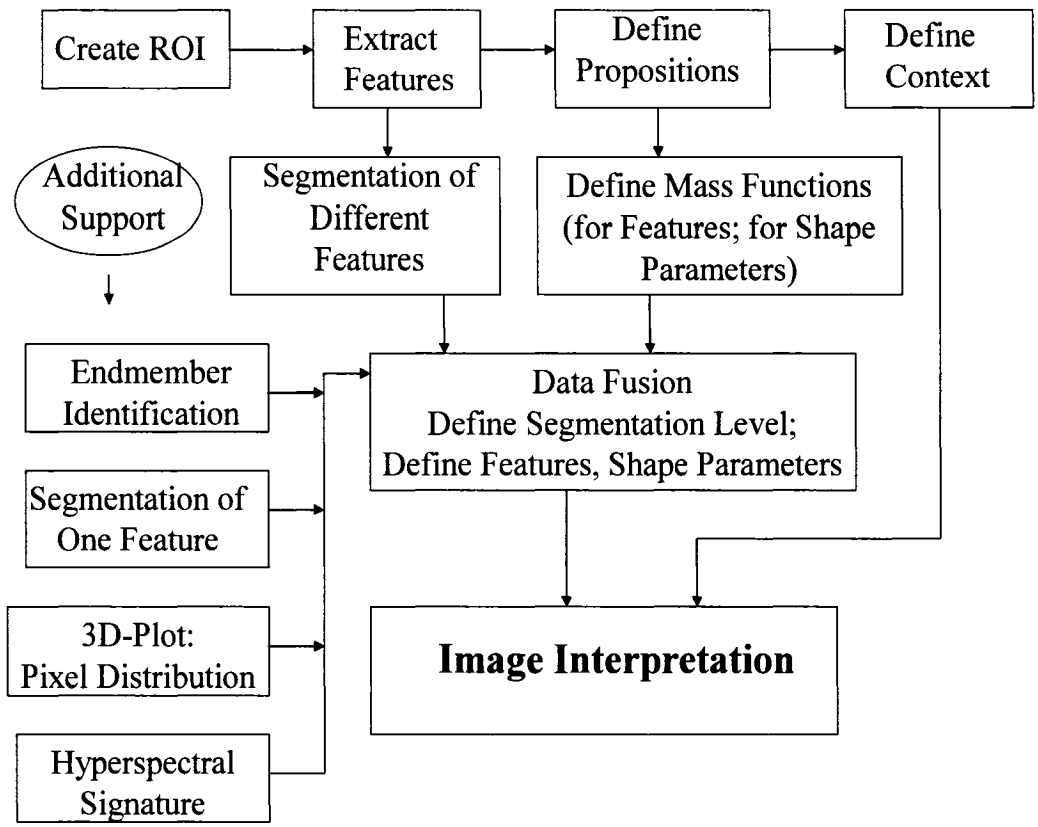


Figure D-2: Task hierarchy for hyperspectral image interpretation

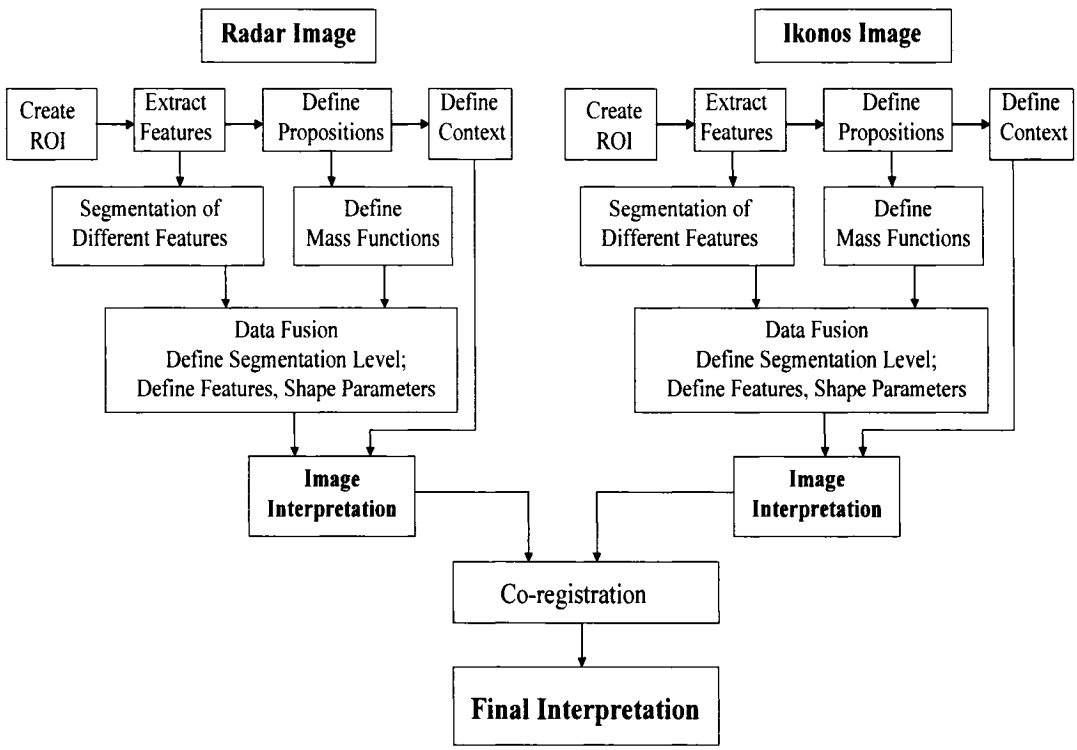


Figure D-3: Task hierarchy for radar/ikonos fused image interpretation

APPENDIX E

USABILITY QUESTIONNAIRE PROPOSED BY ROBERT

1. Facteurs d'utilisabilité des interfaces humains-ordinateur

Le tableau 1 présente les critères d'évaluation de l'utilisabilité IHO (interfaces humains-ordinateur) qui sont proposés par trois différents groupes d'auteurs. Malgré quelques divergences de terminologie, d'ordre ou de regroupement, les auteurs s'entendent sur l'essentiel.

Tableau E.1 : Critères d'évaluation de l'utilisabilité des IHO

Bastien (1991)	Ravden & Johnson (1989)	Robert (1990)
1. Guidage 2. 2. Charge de travail 3. Contrôle explicite 4. Adaptabilité 5. Gestion des erreurs 6. Homogénéité/ consistence 7. Signifiance des codes et dénomination 8. Compatibilité	1. Clarté visuelle 2. Cohérence 3. Compatibilité 4. Bon retour d'information 5. Caractère explicite **6.** Fonctionnalités appropriées **7.** Flexibilité et contrôle **8.** Prévention et correction des erreurs **9.** Guidage et soutien à l'utilisateur	**Fonctionnalités appropriées pour la tâche** **Qualités de l'interface :** -Compatibilité -Transparence -Cohérence -Contrôle explicite -Rapidité d'actions - Flexibilité/ adaptation **Fonctionnalités-clés pour l'interaction** -Retour d'information -Navigation -Aide -Gestion des erreurs

Les paragraphes qui suivent définissent ces facteurs d'utilisabilité et proposent une série de questions ou de lignes directrices pour aider les concepteurs. Plusieurs questions sont tirées de Ravden & Johnson (1989) et ont été traduites, adaptées et complétées. Noter que certaines questions pourraient être classées sous plusieurs facteurs.

1.1 Fonctionnalités appropriées pour la tâche

Une interface doit d'abord et avant tout permettre de faire la tâche pour laquelle elle existe. Elle doit ainsi comporter les fonctionnalisés qui sont requises par la tâche. Offrir le bon niveau de fonctionnalisés dans une interface représente un défi important pour le concepteur.

1. Est-ce que le système comprend toutes les fonctionnalités requises pour faire la tâche ?
2. Est-ce que le système comprend les seules fonctionnalités requises pour faire la tâche ?
3. Est-ce que les fonctionnalités du système sont appropriées pour faire la tâche ?
4. Est-ce que chaque écran contient l'information requise pour faire une partie de la tâche ?
5. Est-ce que le mode de présentation de l'information à l'écran est approprié à la tâche ? (ex.,formulaire, tableau, bloc d'information)
6. Est-ce que l'utilisateur a accès à toute l'information qui est requise pour faire la tâche ?
7. Est-ce que le système permet à l'utilisateur de faire tout ce qu'il juge nécessaire pour réaliser la tâche ?
8. Est-ce que le jargon et la terminologie qui sont utilisés dans le système sont définis au début de la tâche ?
9. Est-ce que les dispositifs d'entrée de données du système sont appropriés pour faire la tâche ? (ex., clavier, souris, bâton de commande, ...)

10. Là où l'interface du système utilise une métaphore (ex., de bureau), est-ce que celle-ci est appropriée pour la tâche ?

1.2 Qualités des IHO

Les qualités que nous présentons dans cette section s'appliquent à l'ensemble des composantes des IHO.

1.2.1 Compatibilité

La compatibilité est une notion capitale en ergonomie. Elle peut être définie comme la qualité d'un système qui respecte les conventions établies ou les attentes d'une population donnée dans un domaine défini. Cette notion comporte deux facettes: la compatibilité par rapport à la tâche qui consiste à respecter la logique, la structure, l'organisation ou la terminologie de la tâche, et la compatibilité par rapport aux utilisateurs qui consiste à respecter les stéréotypes ou les attentes des utilisateurs à l'égard de divers concepts (ex., la couleur rouge, signe de tête affirmatif) ou objets (ex., un bouton poussoir, un interrupteur). Cette notion est directement liée aux stéréotypes d'une population donnée. Attention: la compatibilité porte sur des aspects liés à la culture d'une population.

1. Est-ce que la séquence des activités requises pour faire la tâche (ex., les procédures) correspond aux attentes des utilisateurs?
2. Est-ce que l'information affichée à l'écran correspond à la perception que les utilisateurs ont de la tâche?
3. Est-ce que les graphiques présentent bien le type d'information auquel les utilisateurs
4. s'attendent? (ex., évolution de la température -degré Celsius- en fonction du temps -min-)

5. Est-ce que l'organisation et la structure du système correspondent à la perception que les utilisateurs ont de la tâche?
6. Est-ce que les couleurs utilisées correspondent aux conventions établies là où elles existent (ex., rouge = arrêt; vert = autorisation)?
7. Est-ce que le jargon et la terminologie utilisés sont familiers à l'utilisateur ?
8. Est-ce que les abréviations, les acronymes, les codes et autres informations alphanumériques qui sont utilisés
    - sont faciles à reconnaître et à comprendre?
    - correspondent aux conventions établies là où elles existent?
9. Est-ce que les symbole:" les icones, les représentations graphiques ou picturales qui sont utilisés
    - sont faciles à reconnaître et à comprendre?
    - correspondent aux conventions établies là où elles existent?
10. Est-ce que les unités de mesure utilisées dans le système (ex., minutes, mètres cube, kg, dollars, degré Celsius) correspondent
    - aux attentes des utilisateurs ?
    - aux conventions établies là où elles existent?
11. Est-ce que le fonnat de présentation des informations (ex., date, heure, adresse, numéros, montants ...) correspond
    - aux attentes des utilisateurs?
    - aux conventions établies là où elles existent?
12. Est-ce que le format de présentation de 1'infonnation affichée à l'écran est conforme à celui des données à entrer?
13. Est-ce que le format des écrans est compatible avec celui des documents papier?
14. Est-ce que la relation spatiale entre les dispositifs de présentation d'informations (ex., voyant lumineux, cadran, écran, ...) et les commandes correspond aux attentes des utilisateurs?

15. Est-ce que la direction des actions de l'utilisateur sur un objet (ex., déplacer la souris vers la droite) correspond à la direction du mouvement de l'objet (ex., le curseur se déplace vers la droite)?
16. Est-ce que le format et la séquence de présentation de l'information qui est imprimée sont les mêmes que ceux de l'information qui est affichée à l'écran?

1.2.2 Cohérence

C'est la qualité d'une interface qui respecte les règles qui régissent la signification, l'organisation, la représentation ou le comportement de ses diverses composantes (ex., les fenêtres se manipulent toujours de la même manière dans le système). Cette notion est proche de l'homogénéité.

1. Est-ce que le système est cohérent avec les autres systèmes que l'utilisateur manipule? (ex., même disposition des touches sur le clavier, mêmes procédures de travail)
2. Est-ce les différentes couleurs sont utilisées de façon cohérente à travers le système?
3. Est-ce que la terminologie et le jargon relatifs à la tâche et au système sont utilisés de façon cohérente à travers le système?
4. Est-ce que les abréviations, les"acronymes, les codes et autres informations alphanumériques sont utilisés de façon cohérente à travers le système?
5. Est-ce que les icones, les symboles, les représentations graphiques ou picturales sont utilisés de façon cohérente à travers le système ?
6. Est-ce que les mêmes types d'informations (ex., titres, menus, messages, instructions, ...) sont toujours présentés
   - au même endroit de l'écran ?
   - dans le même format ?
7. Est-ce que les procédures à suivre pour réaliser des tâches de même type sont cohérentes à travers le système?

8. Est-ce que le curseur occupe la même position initiale sur des écrans de type semblable?
9. Est-ce que l'action requise pour déplacer le curseur sur l'écran est cohérente à travers le système?
10. Est-ce que la méthode de saisie de données est cohérente à travers le système ?
11. Est -ce que le format dans lequel I ' utilisateur entre certains types de données à l'écran est le même à travers le système?
12. Est-ce que la méthode de sélection utilisée (ex., pour les éléments d'un menu, une partie de texte) est cohérente à travers le système?
13. Là où on utilise un clavier, est-ce que les mêmes touches correspondent aux mêmes fonctions à travers le système ?(ex., raccourcis-clavier, touches-fonctions)
14. Est-ce que le système répond toujours de façon cohérente aux actions spécifiques de l'utilisateur?

1.2.3 Clarté visuelle

Selon ce facteur, l'information affichée à l'écran devrait être claire, bien organisée, sans ambiguïté et facile à lire.

1. Est-ce que chaque écran est clairement identifié par un titre significatif?
2. Est-ce que l'information qui est importante à l'écran (ex., curseur, instruction, message
3. d'erreur) est bien mise en relief?
4. Lorsqu'il y a des données à entrer dans le système, est-ce que l'utilisateur sait :
   - où les entrer à l'écran?
   - dans quel format les entrer?
5. Est-ce que l'information affichée à l'écran est facile à voir et à lire?
6. Est-ce que les informations sont bien disposées à l'écran?
7. Est-ce que les informations semblent organisées de façon logique à l'écran? (ex., ordre des menus, ordre des éléments des menus)

8. Est-ce que l'utilisation de la couleur aide à présenter clairement l'information à l'écran?
9. Est-ce que le code des couleurs utilisées dans le système est facile à découvrir et comprendre?
10. Est-ce que l'information qui est affichée en couleurs est facile à voir même si l'écran a une mauvaise résolution graphique ou pour les utilisateurs daltoniens?
11. Lorsque l'utilisateur entre des données qui vont en remplacer d'autres, est-ce que le système élimine les données existantes pour éviter toute confusion entre les deux groupes de données?
12. Est-ce que les différents types d'informations présentés à l'écran sont clairement séparés sur l'écran? (ex., étiquettes, champs de saisie, instructions)
13. Lorsqu'il y a beaucoup d'information affichée à l'écran, est-ce que celle-ci est clairement séparée en sections?
14. Est-ce que les colonnes d'informations sont bien alignées à l'écran? (ex., texte alphanumérique justifié à gauche, chiffres justifiés à droite)
15. Est-ce que des couleurs claires ou brillantes sont présentées sur des fonds foncés à l'écran, ou vice versa?
16. Est-ce qu'il est facile de trouver l'information que l'on recherche à l'écran? (ex., les fonctionnalités offertes par le système)
17. Est-ce que les représentations graphiques et picturales (ex., figures, diagrammes, images) ont bien dessinées et annotées?

1.2.4 Caractère explicite

Le caractère explicite est la qualité d'une interface dans laquelle le mode de fonctionnement et la structure du système sont clairs pour l'utilisateur. Cette notion est très proche de la notion de transparence.

1. Est-ce que l'état ou le mode d'opération dans lequel se trouve le système à chaque étape de la tâche est clair pour l'utilisateur?

2. Est-ce que la structure du système est claire pour l'utilisateur?
3. Est-ce que la raison pour laquelle le système est structuré et organisé ainsi est claire pour l'utilisateur?
4. Est-ce que ce qui doit être fait pour réaliser la tâche est clair pour l'utilisateur?
5. Est-ce que la signification de chaque option offerte par le système (ex., dans un menu) est claire pour l'utilisateur?
6. Est-ce que la raison pour laquelle des changements effectués dans une partie du système affectent d'autres parties du système est claire pour les utilisateurs? Est-ce que l'endroit où ces changements se produisent est clair pour les utilisateurs ?
7. Est-ce que la partie du système dans laquelle l'utilisateur se trouve à un instant donné est claire pour ce dernier?
8. Est-ce que le rôle des différentes parties du système est clair pour les utilisateurs?
9. Si le système utilise une métaphore pour l'interface (ex., métaphore de bureau dans les logiciels de bureautique), est-ce que celle-ci est explicite?
10. Si la métaphore de l'interface ne s'applique qu'à certaines parties du système, est-ce que cela est explicite?

1.2.5 Flexibilité/ Adaptation

C'est la qualité d'une interface qui offre des choix à l'utilisateur quant à la façon de faire des opérations, qui permet de modifier l'environnement de travail selon ses goûts ou préférences, ou qui s'adapte aux besoins et modes de fonctionnement des utilisateurs.

1. Est-ce que le système offre un choix de dispositifs d'entrée de données ou de pointage pour travailler? (ex., clavier vs souris)
2. Est-ce que le système offre un choix de procédures pour faire certaines opérations? (ex., pouvoir sélectionner du texte de n manières différentes)
3. Est-ce que l'utilisateur peut modifier certains aspects de l'interface selon ses besoins ou ses préférences ? (ex., couleurs, vitesse de la souris, forme du curseur)

4. Est-ce que l'utilisateur peut définir le nom et l'organisation de l'information à laquelle il veut avoir accès ultérieurement ? (ex., fichier, dossier)
5. Est-ce que le système donne accès à différents niveaux d'aide en-ligne pour accommoder différents utilisateurs ?
6. Est-ce que l'utilisateur a le choix d'entrer l'information à la main ou de laisser le système la produire automatiquement ? (ex., là où il y a des entrées répétitives de données et des valeurs par défaut)
7. Est-ce que l'utilisateur peut choisir l'ordre dans lequel il fait des requêtes d'informations au système ou celui dans lequel il effectue des activités?

1.2.6 Rapidité d'action

C'est la qualité d'une interface qui permet à l'utilisateur de faire des opérations avec le minimum d'actions et de travailler rapidement. Ce facteur est surtout important pour les opérations fréquentes ou urgentes, ou lorsque la charge de travail est élevée et que le temps presse. Elle est proche des notions de concision et de charge de travail.

1. Est-ce que le système utilise les procédures les plus brèves possible? (surtout pour les opérations courantes)
2. Est-ce que le système permet à l'utilisateur de définir des macro-commandes?
3. Est-ce que le système permet d'utiliser des raccourcis lorsque c'est requis? (ex., dans les menus, pour éviter une séquence d'activités ou d'écrans)
4. Dans les systèmes basés sur des menus, est-ce que l'utilisateur peut revenir rapidement au menu de départ? (ex., par une seule touche)
5. Est-ce que le système permet d'utiliser des valeurs par défaut dans certains champs de saisie?
6. Est-ce que le système remplit à l'avance certains champs de saisie dans les cas d'entrée répétée de données?
7. Est-ce que le système permet de sélectionner des données dans un menu plutôt que de les taper lorsqu'elles sont pré-définies? (ex., boîte combo)

8. Est-ce que le système permet à l'utilisateur de sortir d'une procédure sans devoir la compléter ou d'arrêter des opérations en cours d'exécution ? (ex., impression, calcul)
9. Est-ce que le système permet d'accéder directement à un écran particulier faisant partie d'une série d'écrans ?
10. Est-ce que le système permet d'aller rapidement à différents endroits d'un document ? (ex, Début, Fin, page n, endroits définis par des marqueurs)
11. Est-ce que le système permet de faire des aller-retours rapides entre l'état actuel et l'état précédent?

1.3 Fonctionnalisés-clés pour l'interaction

Quatre types de fonctionnalisés sont particulièrement importantes pour l'utilisabilité d'une IHO: le retour d'information, l'aide/le guidage, la gestion des erreurs et la navigation.

1.3.1 Retour d'information

Le retour d'information correspond à toute forme de réponse du système faisant suite à une action de l'utilisateur, ou a toute information provenant du système pour informer l'utilisateur du statut ou mode du système. Il peut être visuel, auditif, les deux à la fois, ou tactile (ex., blocage du clavier).

1. Fournir un retour d'information pour chaque action de l'utilisateur, qu'elle soit valide ou non (ex., entrée de données, sélection d'un élément dans un menu)
2. Le retour d'information doit être rapide, significatif et présenté là où l'action se déroule. (< 2 sec pour les opérations normales; durée compatible avec le type de transactions en cours).
3. Informer l'utilisateur du statut de la transaction lorsque le temps de réponse va être plus long.

4. Informer clairement l'utilisateur lorsque le système a terminé de traiter sa requête.
5. Si l'utilisateur demande de détruire des objets qui sont stockés dans l'ordinateur mais qui ne sont pas affichés à l'écran, lui montrer ces objets afin qu'il confirme ou annule la commande avant que la transaction soit complétée.
6. Indiquer clairement à l'utilisateur quelles touches-fonctions ou quels éléments des menus sont actifs et lesquels ne le sont pas.
7. Lorsque l'utilisateur utilise une imprimante qui est éloignée de son poste de travail, lui présenter un message pour l'informer que sa demande d'impression est en traitement: soit en préparation, en attente (queue), ou en cours d'exécution.
8. Si l'utilisateur interrompt un traitement qui est en cours (ex., calcul, impression), lui présenter un message pour l'informer que le système est revenu à son statut précédent.
9. Fournir automatiquement un retour d'information lors de transmission de données, pour confirmer que le message a été envoyé ou indiquer qu'il n'y a pas eu de transmission afin que l'utilisateur puisse s'ajuster en conséquence.
10. Dans les cas de non transmission de messages, si possible expliquer la cause du problème.

1.3.2 Aide/Guidage

L'aide correspond à toute forme d'assistance fournie à l'utilisateur pour le conseiller, le guider, l'informer, et le conduire lors de ses interactions avec le système. Elle comprend l'aide en ligne, les différents types de messages, les instructions, les incitations, les cartes d'aide, les bulles d'information, etc.

1. Offrir de l'aide en ligne au moyen de la fonction AIDE ou?
2. Permettre d'entrer facilement dans la fonction Aide au moyen d'une action simple et standard (ex., cliquer sur le?).

3. Permettre de se déplacer facilement dans le contenu de l'aide et d'en sortir rapidement.
4. Adapter l'aide au contexte de la tâche et à la transaction en cours.
5. Donner accès à l'aide à partir de n'importe quel point dans le système. I'-.
6. Permettre de consulter rapidement les différents sujets traités dans la fonction Aide et de faire un choix.
7. Offrir différents niveaux d'aide pour répondre aux besoins de différents utilisateurs.
8. Fournir des informations sur le format des données valides dans certains champs de saisie (ex., Date: "„ (jour-mois-année».
9. Guider l'utilisateur dans l'exécution des procédures (ex., donner des instructions).
10. Présenter les messages et les instructions dans un style clair, concis et positif.
11. Présenter des instructions et des incitations (prompt) qui indiquent clairement ce que l'utilisateur doit faire.

1.3.3 Gestion des erreurs

Les fonctionnalités de prévention et de correction des erreurs, incluant la détection, l'explication et le recouvrement, sont essentielles dans tous les systèmes interactifs, parce que les erreurs sont inévitables, quel que soit notre niveau de qualification avec le système, ou la qualité du système que nous utilisons.

1. Bloquer l'entrée de donnée non valides provenant de l'utilisateur afin de prévenir des erreurs ou des pertes.
2. Informer clairement et rapidement l'utilisateur dès qu'une erreur a été détectée par le système.
3. Permettre à l'utilisateur de revenir facilement en arrière avec la fonction Cancel pour annuler des actions en cours de définition et la fonction Undo pour défaire le résultat des traitements.
4. Permettre à l'utilisateur d'arrêter des traitements en cours d'exécution (avec la fonction Stop).

5. Demander à l'utilisateur de confirmer des requêtes ou des actions ayant des conséquences graves ou qui ne peuvent pas être défaites facilement (ex., sortir du système alors qu'il y a des transactions en cours et que des données peuvent être perdues).
6. Présenter des messages d'erreurs spécifiques et informatifs, qui expliquent clairement la nature, la localisation et la cause de l'erreur, et comment faire pour la corriger.
7. Adopter un ton neutre dans les messages d'erreurs: éviter de blamer l'utilisateur, ou de personnaliser le système (ex., 'J'ai besoin d'un nombre entier"), ou de présenter des messages comprenant de l'humour.
8. Permettre à l'utilisateur de demander des explications plus détaillées sur une erreur.
9. Adopter un langage approprié à la tâche dans les messages d'erreurs.
10. Présenter un message d'erreur seulement après que l'utilisateur ait fini d'entrer les données.
11. Enlever le message d'erreur une fois l'erreur corrigée.
12. Lorsque des erreurs de système se produisent, donner accès à toute l'information nécessaire pour faire le diagnostic et résoudre le problème.

1.3.4 Navigation

Les fonctionnalités de navigation permettent de se déplacer dans l'interface; comme elles sont très fréquemment utilisées, leur rôle est crucial pour l'utilisabilité du système. Elles comprennent les touches de déplacement du curseur sur le clavier, la touche de retour de chariot, la touche du tabulateur, les touches de positionnement du curseur (ex., début, fin d'une ligne, de la page, du document), les commandes pour aller directement à une certaine page du document, les signets, la boîte d'ascenseur à l'écran, etc.

1. Placer le curseur à un endroit utile et cohérent dans tout nouvel écran.
2. Permettre de se déplacer facilement et rapidement à tout endroit d'un document.

3. Passer facilement d'un champ de saisie à l'autre à l'écran (ex., en utilisant la touche du tabulateur, le retour de chariot, la souris).
4. Passer directement d'une colonne à l'autre, ou d'une ligne à l'autre dans une même colonne dans des tableaux.
5. Permettre de définir et de positionner des signets afin d'accéder rapidement à des points particuliers du document.